

QY	9930	AATGATGATGTCGCGCATATGAGC-----TCTCACTGATGTCCTGGCTGTAACCTGGCTGCT	983
	Db	343 AmsapMeLeuValValAsnGlnGlnAlaAsnLeuLeuProGlyLeuAsnLeuSerAla	362
QY	984	GTAGGTCTTTTCCAGAGCTTCATCCAGCGAGTCGCCGCGCT-----CCAGAGAGCGCT	103
Db	363	LeuGlyIlePheSerThrGlyLeuSerValIleuProProAlaGlyProArgIlyVal	382
QY	1038	ACTGGAGCTGCTCCCTATAGCTCCCTTANG-----	106
Db	383	ProProSerProProTyrHisProPheAlaThrHisSerGlyTyrPheSerSerLeuTyr	402
QY	1068	-----CAGGCTCCGAGCAGAGATGATG	1091
Db	403	ProHisHisHisPheGlyProPheProHisHisHisSerTyrProGlyGlnIleThrVal	422
QY	1092	CAGGTGTTATCCCGCCAGCAGTGGAGCGGCATCATCGGAGAAAGAGGAGACATC	115
Db	423	SerLeuPheIleProThrGlnAlaValGlyAlaIleIleGlyLysGlyAlaHisIle	442
QY	1152	AAAGAGCTCTCCGAGTTTGCCAGCGCCTCCATCAGATTGGACACCCGAAACACTGAC	1211
Db	443	LysGlnLeuAlaArgPheAlaGlyAlaSerIleLysIleAlaProAlaGlnIlyProAsp	462
QY	1212	TCCAAAGTTCGTATGTTATATCATCACTGGACCGCCAGCGCCCATTAAGGCTCAGGGA	1271
Db	463	ValSerGluArgMetValIleIleThrGlyProProGluAlaGlnPheLysValSerAla	482
QY	1272	AGA 1274	
Db	483	Lys 483	
RESULT 15			
Q9VZ69	PRELIMINARY;	PRT;	566 AA.
AC	Q9VZ69		
DT	01-MAY-2000 (TrEMBLrel. 13, created)		
DT	01-OCT-2002 (TrEMBLrel. 22, last annotation update)		
DT	01-OCT-2003 (TrEMBLrel. 25, last annotation update)		
DE	CG1691 protein (IGF-II mRNA-binding protein) (SD07045p).		
OS	IMP OR CG1691.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pezomyzeta;		
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Berkely.		
RX	MEDLINE=20196006; PubMed=10731132;		
RA	Adams M.D., Celisniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,		
RA	Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,		
RA	Avril J.F., Agbayani A., An H.-D., Andrews-Pfannkoch C., Baldwin D.,		
RA	Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu I., Beasley E.M.,		
RA	Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,		
RA	Borkova D., Borchan M.R., Bouck J., Brokstein P., Brotlier P.,		
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,		
RA	Dodson K.J., Evansglista C.C., Ferraz C., Ferreira S., Fleischmann W.,		
RA	Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,		
RA	Fosler A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,		
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Idegiam C.,		
RA	Jalali M., Kalush F., Karpem G.H., Ke Z., Kemison J.A., Ketchum K.A.,		
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,		
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,		
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,		

RA Merkurov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Murzy D.K., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paclet J.M.,
RA Palczolo M., Peltington K., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Sanders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskys R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein C.W.M., Weissbach U.
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye Y., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*";
RL Science 287:2185-2195(2000).
RN [2]
RN
RN
RN SEQUENCE FROM N.A.
RP Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Barton J.J., An H., Baldwin D., Banazon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Gary N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Idagwam C., Jalili M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunco J.,
RA Paclet J., Parasas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svitskys R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of *Drosophila melanogaster* genome";
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RN
RN
RN SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hrdecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Rudy J.L., Bergman C., Bertram B., Carlisle J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn B., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of *Drosophila melanogaster* genome";
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [4]
RN
RN
RN SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [5]
RN
RN
RN SEQUENCE FROM N.A.
RA Flybase;
RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
RN [6]
RN
RN
RN SEQUENCE FROM N.A.
RA Nielsen J., Nielsen F.C., Christiansen J.;
RT "Cloning and Expression of a *Drosophila* Homolog of IMP/VgI-RBP";
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [7]
RN
RN
RN SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Abpayani A., Carlson J.,
RA Champes M., Chavez C., Dorsett V., Fatian D., Frise E., George R.,
RA Gonzalez M., Guarini H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunco J., Paclet J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
CC -I- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR EMBL; AE003484; AAF47958.2; -
DR EMBL; AF241237; AAF63331.1; -
DR EMBL; AY069821; AAL39966.1; -
DR FlyBase; FBgn0030235; Imp.
DR GO; GO:0003676; Ribonucleic acid binding; IEA.
DR InterPro; IPRO04087; KH_dom.
DR InterPro; IPRO04088; KH_type_1.
DR Pfam; PF00013; KH; 4.

```
DR SMART; SM00322; KH; 4.
DR PROSITE; PS50084; KH_TYPE_1; 4.
DR SEQUENCE 566 AA; 62129_MW; B4C74C5C99B0C830 CRC64; 30
```

Alignment Scores:

Pred. No.:	2,678-55	Length:	566
Score:	863.50	Matches:	203
Percent Similarity:	56.60%	Conservative:	93
Best Local Similarity:	38.81%	Mismatches:	150
Query Match:	24.45%	Indels:	75
DB:	5	Gaps:	12

US-09-270-437D-7 (1-1946) X Q9VZ69 (1-566)

QY	213	AAACCGAAGTAGAGACGGAGTGGTGAATATGTCACCTATTCCCAACGGAGACGACG	272
Db	4	Asnsmnsmberserhraglleuasnsmnllseer---Asnsmnlyrlyrglnghlnlys	22
QY	273	CAACCATCATGAG-----CTGAATGCCACCAAGTTGGAG	308
Db	23	Glnserneuileargtrylleuaspargalalavalglyleuasnghlyalghnphegln	42
QY	309	AACCATGCCCTGAAGGCTCCTCATCCCGAGTAGACAGATACACAGGACCTGAGAT	368
Db	43	Glyserlyseuhtisalaglughlnleuaspulyasngln-----	55
QY	369	GGGCGCGAGGGGGCTTTGGCTCGGGGATGACCCCGACGGGCTCACCTGTGCAAGC	428
Db	56	-----Argargserghlnargsmghlnargasnprotyr---Pro	67
QY	429	GGGGCCCCACGCAAGACGACGACGACATCCCCCTTGGGCGCTGGGGCCACCGAG	488
Db	68	GlymetProglyProglyArgghlnlaaspheProleuarglleuvalghnserghl	87
QY	489	TATGGGGTGCATTTATGTCAGAGAGGGGGCCACATCCGCAACATCACAAAACGACC	548
Db	88	Metvalglyalaitlelleghlyargghnghlyserthrilleargthrillethrglnghnser	107
QY	549	CAGTCCAAAGATAGACGTGCATAGAGAGAGAACCGCAGGTGCAGCTGAAAAGCCATCAGT	608
Db	108	Argalaargvalasplvalhlsarglysglnasnvalglyserleuughlyussettlethr	127
QY	609	GTGCACTCCACCCCGAGAGCGTGCCTCCGCTGTGAAGATGATCTTGAGATTATGAT	668
Db	128	lletryglyasnproghlyasnscythrasmnlaayshysargilleuughlnvalmetghln	147
QY	669	AAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCTCGAAGATCTGGCCCATAT	728
Db	148	ghnghlnlaitleserThrasmlysegly---ghllecyseuhyisleuvalhlsasn	166
QY	729	AACCTTGAGGGCGTCTCATTTGGCAAGAGACGCAACCTGAGAAAGGTAGACAGAT	788
Db	167	Asnleuilelleghlyargllelleghlyssserghlysnthrilleysargillemetghlnsp	186
QY	789	ACCGAGCAAAAATACACCATCTCTCGTGTGCAAGACCTTACCTTTACACCTGAGAG	848
Db	187	ThraspThrlyllethrValserserilleasnspilleasnserPheasnleughlyarg	206
QY	849	ACCATCATCTGAGAGGGGGCCATCCAGAAATTGTTCGACAGGGCGACGAGAAATATAGAG	908
Db	207	llellethrVallyseghlyleuileghlyasnmetserltharglalnghlnsnghlnlethr	226
QY	909	AAAGTTCCGGAGAGCCATGAGATGATGTGGCTGCATGAGC-----TTCACCTGATC	962
Db	227	Lysleuargghnsertryghlnsmnspheughlnlmetalaproghnserleuemetp	246
QY	963	CCTGCCCTGAACCTGGCTGCTGAGGTCTTTTCCAGCTTCA-----TTCAGC	1010
Db	247	ProghlyleuhtisPrometalametmetserThrProghlyasnghlymetvalPheasnthr	266
QY	1011	GCAGTCCGCGCGCT-----CCAGAGCGGTACTCGGGGCTGGCTCCCAT	1055
Db	267	SerMetProhpheProserCysghnserPhealmetserlysrthProhalservalval	286

OY	1056	AGCTCTTTATGCAGAGGTCCCGACGAGATGCTGACGGTGTATACCCCGCCGACGCA	1115
Db	287	ProProValAlpheProAsnAspLeuGlnGlnThrThryLeuTyxTlleProAsnAsnAla	306
OY	1116	GTGGGCGCCATCATCGGCGAAGAGGGGACACATCAACAAGCTCTCCGGTTTGCCAGC	1175
Db	307	ValGlyAlaTlleTleGlyThrArgGlySerHisTlleArgSerTlleMetArgPheSerAsn	326
OY	1176	GCTTCATCAAGATMTGACCACCCGAAACA-----CCTGACTCCAAAGTTGCT	1223
Db	327	AlaSerLeuTyxTlleAlaProLeuAspAlaAspLysProLeuAspGlnGlnThnGlnArg	346
OY	1224	ATGGTTATCATCATGTGACCGCCGACAGGCCCAATTCAGGCTCAGGGAAAGATATATGGC	1283
Db	347	LysValAlhTrIleValGlyThrProGlnGlyGlnTrpLysAlaGlnTyMetIlePheGlu	366
OY	1284	AAACTCAAGAGAGAGAACTCTTGGTCCCAAGAGAGAGAGAGAGTGAAGTGTGGAGACCCACATA	1343
Db	367	LysMetArgGlnGlnGlyPheMetCysGlyThraspAspValArgLeuThnValGlnLeu	386
OY	1344	CGTGTGCAGCATCAGCAGCTGCGCGGGGTCAATTGGCAAAAGGTGAAAAACGTGAACAG	1403
Db	387	LeuValAlaSerSerGlnValGlyArgGllleGlyLysGlyGlnAsnValArgGlu	406
OY	1404	TTGCAGATTGTGACGGCAGCTGAGGTGGTATACCAGA-----	1442
Db	407	LeuGlnArgValThrGlySerValTlleLysLeuProGlnHisAlaLeuAlaProProSer	426
OY	1443	-----GACAGACCCCTGATGAGAACGACACAGCATCTGAAAAATCATGGACAT	1493
Db	427	GlyGlyAspGlnGlnThnPro-----ValHisTlleTlleGlyLeu	439
OY	1494	TTTCTATGCCAGTCAGATGGCTCTCAACGGAAGATCCGAGACATCCGGGCCACG-----	1544
Db	440	PheTyxSerValGlnSerAlaGlnArgTlleArgAlaMetMetLeuSerThrAsnPro	459
OY	1545	-----GTTAAGCAGCAGCATCAGAACGAGCAAGATAC	1577
Db	460	ProProTlleThnTyLysGlnLysAlaAlaLysGlnGlnLeuGlnGlnGlnGlnIns	479
OY	1578	CAGGCCCGACGACGAGAGAGTGCACCAAGCCCTCCCTGTCCCTTNGATCCAGGACACA	163
Db	480	LeuAlaGlyAla-AlaSerSerGlySerGlnGlnGlnGlnProGlnSerProSerGlnGln	499
OY	1638	ACGGACA 1644	
Db	499	nglnAla 501	

Search completed: July 23, 2004, 11:15:00
Job time : 120.039 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2004, 03:38:17 ; Search time 12641.7 Seconds

(without alignments)
11256.058 Million cell updates/sec

Title: US-09-270-437D-8

Perfect score: 3283

Sequence: 1 ggcagcgagagagcgagga.....aaccttgaaatgttattc 3283

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 segs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_hcg_hum.*
31: em_hcg_inv.*
32: em_hcg_other.*
33: em_hcg_mus.*
34: em_hcg_pln.*
35: em_hcg_rod.*
36: em_hcg_mam.*
37: em_hcg_vrt.*
38: em_sy.*
39: em_hcgo_hum.*
40: em_hcgo_mus.*
41: em_hcgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3282	100.0	3283	6	AR171867
2	3282	100.0	3283	6	AR343076
3	3282	100.0	3283	6	BD209927
4	3253.8	99.1	3667	9	AF057352
5	3143	95.7	3412	6	AR171865
6	3143	95.7	3412	6	AR343074
7	3143	95.7	3412	6	BD209925
8	2799.4	85.3	142971	9	AC020629
9	2781.4	84.7	91084	9	AF596177
10	2779.8	84.7	182695	2	AC015706
11	2327	70.9	187226	9	AC104980
12	2325.4	58.6	3557	10	BC023758
13	1924.2	52.1	2010	9	AF117107
14	1711.8	49.3	2640	9	AC016961
15	1627	49.6	98945	9	HSM806243
16	1613	49.1	208273	2	AC134942
17	993.6	30.3	4953	10	BC054552
18	748.6	22.4	156398	2	AC094924
19	734.8	22.4	2021	5	AF026527
20	718.2	21.9	2105	10	AF541940
21	707.2	21.5	50320	2	AC126373
22	700.6	21.3	2130	9	AF117106
23	697.2	21.2	2381	9	AF198254
24	697.2	21.2	1740	6	BD275902
25	697.2	21.2	1740	6	AR220687
26	697.2	21.2	1740	6	AR281251
27	697.2	21.2	1740	6	AX365954
28	697.2	21.2	1740	6	AX366054
29	697.2	21.2	1740	6	AX366057
30	697.2	21.2	1740	6	AX366057
31	697.2	21.2	1740	6	AX366057
32	695.6	21.2	4159	6	AR171863
33	695.6	21.2	4159	6	BD209923
34	695.6	21.2	4159	6	BD209923
35	695.6	21.2	4159	6	BD209923
36	695.6	21.2	4159	6	BD209923
37	695.6	21.2	4159	6	BD209923
38	695.6	21.2	4159	6	BD209923
39	695.6	21.2	4159	6	BD209923
40	695.6	21.2	4159	6	BD209923
41	695.6	21.2	4159	6	BD209923
42	695.6	21.2	4159	6	BD209923
43	695.6	21.2	4159	6	BD209923
44	695.6	21.2	4159	6	BD209923
45	695.6	21.2	4159	6	BD209923

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
AR171867	AR171867	Sequence 8 from patent US 6297364.	AR171867.1	GI:17910817	Unknown.	Unknown.	1 (bases 1 to 3283)	Chen,Y.-T., Gure,A., Tsang,S., Stockert,E., Jager,E., Alexander,K. and Old,L.J.	Isolated nucleic acid molecule encoding cancer associated antigen, the antigen itself, and uses thereof

QY 2041 GCGTATCCCTTTTAACTGTAACATAGTGAGCGTGTTCAAAGCCAAAGATGAC 2100
Db 2041 GCGTATCCCTTTTAACTGTAACATAGTGAGCGTGTTCAAAGCCAAAGATGAC 2100
QY 2101 ACCCTTTTCTGTGCGAAATGCTCTGTGACATGTGTGACATVATTGAAAGGAAATG 2160
Db 2101 ACCCTTTTCTGTGCGAAATGCTCTGTGACATGTGTGACATVATTGAAAGGAAATG 2160
QY 2161 TTAAGATATGTGGCTGTGGGTACACAGGGGCTGACAGGGGTATATATTATTTGAAT 2220
Db 2161 TTAAGATATGTGGCTGTGGGTACACAGGGGCTGACAGGGGTATATATTATTTGAAT 2220
QY 2221 AATATATCAAAATTACTCAACTCAATTTTATCAATTAATTAATTTTCTTTTCTT 2280
Db 2221 AATATATCAAAATTACTCAACTCAATTTTATCAATTTATTAATTTTCTTTTCTT 2280
QY 2281 TTAAGAGAAAGACAGGCTTTTCTGACTTTAAAGATTAAGCTTTGGAGGCTCAGG 2340
Db 2281 TTAAGAGAAAGACAGGCTTTTCTGACTTTAAAGATTAAGCTTTGGAGGCTCAGG 2340
QY 2341 TGTAGAGAGAGCTTTGAGGCGACCCGACAAATTCACCCAGAGGAAATCTGTGGA 2400
Db 2341 TGTAGAGAGAGCTTTGAGGCGACCCGACAAATTCACCCAGAGGAAATCTGTGGA 2400
QY 2401 AGGACTCTCAGGAGTCTGATCAGCTGTGATGTCACAGAGGAGTACCGTCTCT 2460
Db 2401 AGGACTCTCAGGAGTCTGATCAGCTGTGATGTCACAGAGGAGTACCGTCTCT 2460
QY 2461 TGAAGAGAAACTCTGCACTCTCAGCTGTGCTAGCTATACCACTTCTCTTTCG 2520
Db 2461 TGAAGAGAAACTCTGCACTCTCAGCTGTGCTAGCTATACCACTTCTCTTTCG 2520
QY 2521 TTACAGAGTTTAACTGTTTTTGTGATCTGCTATATATATCTCTGTCTCTCTGTT 2580
Db 2521 TTACAGAGTTTAACTGTTTTTGTGATCTGCTATATATATCTCTGTCTCTCTGTT 2580
QY 2581 TATCTCTCCCTCCCTCCCTCCCTCTCTCTCATCTCATTTTGAATTTCCCAT 2640
Db 2581 TATCTCTCCCTCCCTCCCTCCCTCTCTCTCATCTCATTTTGAATTTCCCAT 2640
QY 2641 CCTTCCATCTCAATCCCGTATCTAGCAAGGCGGCGGCGGCGGCGGCGGCGGCGG 2700
Db 2641 CCTTCCATCTCAATCCCGTATCTAGCAAGGCGGCGGCGGCGGCGGCGGCGGCGG 2700
QY 2701 GTATCATATCAACAAAGGAAACAAAGGAAACAAAGGAAACAAAGGAAACAAAG 2760
Db 2701 GTATCATATCAACAAAGGAAACAAAGGAAACAAAGGAAACAAAGGAAACAAAG 2760
QY 2761 GGTACTCAAAAGACAGATGATGTAAGTCTAGCTGTTTGGAGAGGAAACAA 2820
Db 2761 GGTACTCAAAAGACAGATGATGTAAGTCTAGCTGTTTGGAGAGGAAACAA 2820
QY 2821 GGAACCAACCAACCAACCAATCAACCAACCAACCAACCAACCAATCAACCAAT 2880
Db 2821 GGAACCAACCAACCAACCAATCAACCAACCAACCAACCAACCAATCAACCAAT 2880
QY 2881 AATTTGCTTTTGTGATTTGTGTATTAAGCCATCAATATTCAGCAAAATGATTCCTTC 2940
Db 2881 AATTTGCTTTTGTGATTTGTGTATTAAGCCATCAATATTCAGCAAAATGATTCCTTC 2940
QY 2941 TTTAAAAAATATGAGAGAAATAGAAATTTTCAAGGTTGTGGCCAGGGCGTT 3000
Db 2941 TTTAAAAAATATGAGAGAAATAGAAATTTTCAAGGTTGTGGCCAGGGCGTT 3000
QY 3001 AAATTCAGATTTTAAACGAGAAACACACAGAGAGGCTACCTCAGGTTTAA 3060
Db 3001 AAATTCAGATTTTAAACGAGAAACACACAGAGAGGCTACCTCAGGTTTAA 3060
QY 3061 CCTCAGACCTTGTCTGTGTTTCCCTTAGAGATTTTAAAGCTGATGTGGAGCAT 3120
Db 3061 CCTCAGACCTTGTCTGTGTTTCCCTTAGAGATTTTAAAGCTGATGTGGAGCAT 3120

QY 3121 TTTTATTTTTTAAATAAATAGTGGAAAAAATTAAGATATCACTGCCAGCTG 3180
Db 3121 TTTTATTTTTTAAATAAATAGTGGAAAAAATTAAGATATCACTGCCAGCTG 3180
QY 3181 GAGAAGTGACAGTCCAAAGTGTGCAACAGCTGCTGAATGTCTTCGGCTAGCCAAAGA 3240
Db 3181 GAGAAGTGACAGTCCAAAGTGTGCAACAGCTGCTGAATGTCTTCGGCTAGCCAAAGA 3240
QY 3241 CCNATATGCGCTTCTTTTGAACAAACCTTGAATAATGTTATTT 3283
Db 3241 CCNATATGCGCTTCTTTTGAACAAACCTTGAATAATGTTATTT 3283

RESULT 2
AR343076 3283 bp DNA linear PAT 17-AUG-2003
LOCUS AR343076
DEFINITION Sequence 8 from patent US 6576756.
ACCESSION AR343076
VERSION AR343076.1 GI:3738478
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3283)
AUTHORS Chen,Y.-T., Gure,A., Tsang,S., Stockert,E., Jager,E., Alexander,K.
and Old,L.J.
TITLE Isolated nucleic acid molecule encoding cancer associated antigen,
the antigen itself, and uses thereof
PATENT: US 6576756-A 8 10-JUN-2003;
JOURNAL Location/Qualifiers
FEATURES
source 1..3283
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 100.0%; Score 3282; DB 6; Length 3283;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCAGCGGAGAGAGCGAGAGCGCGGCTACCGGCGGAGGCGGCGGCTGTGGGG 60
Db 1 GGCAGCGGAGAGAGCGAGAGCGCGGCTACCGGCGGAGGCGGCGGCTGTGGGG 60
QY 61 AAGAGAGGATGATGAACAAAGCTTTTACATGAGGAACTGAGCCCGCGCTACCGCGGAC 120
Db 61 AAGAGAGGATGATGAACAAAGCTTTTACATGAGGAACTGAGCCCGCGCTACCGCGGAC 120
QY 121 GACCTCCGGAGCTTTTGGGGACAGAAAGCTGCGCTGCGGAGACAGATCTGTGAAG 180
Db 121 GACCTCCGGAGCTTTTGGGGACAGAAAGCTGCGCTGCGGAGACAGATCTGTGAAG 180
QY 181 TCCGGCTAGCGCTTGTGAGCTACCCGACAGAACTGGGCGCATCCGCGCATCGAGAC 240
Db 181 TCCGGCTAGCGCTTGTGAGCTACCCGACAGAACTGGGCGCATCCGCGCATCGAGAC 240
QY 241 CTCTCGGTTAAAGTGAATTCATGAGGAAATCATGAGGAAATGATTAATCAGTCTGAAA 300
Db 241 CTCTCGGTTAAAGTGAATTCATGAGGAAATCATGAGGAAATGATTAATCAGTCTGAAA 300
QY 301 AAGCTAAGAGAGAGAAATTCAGATTCGAAATCATCCCTCTCACTGACGTGGAGGTG 360
Db 301 AAGCTAAGAGAGAGAAATTCAGATTCGAAATCATCCCTCTCACTGACGTGGAGGTG 360
QY 361 TTGATGAGCTTTTGGCTCAATATGAGGACAGTGGAGATGTGGAACAGTCAACAGAC 420
Db 361 TTGATGAGCTTTTGGCTCAATATGAGGACAGTGGAGATGTGGAACAGTCAACAGAC 420
QY 421 AAGAAACCGCGCTTGTCAAGCTCAATATGCAACAGAAAGAAAGCAAAATAGCCATG 480
Db 421 AAGAAACCGCGCTTGTGTCAAGCTCAATATGCAACAGAAAGAAAGCAAAATAGCCATG 480
QY 481 GAGAAGCTAAGGGGATCAGTTTGAAGATCTCTTCAAGATTTCTTACATCCGAGAT 540
Db 481 GAGAAGCTAAGGGGATCAGTTTGAAGATCTCTTCAAGATTTCTTACATCCGAGAT 540

Db 481 GAGAGCTAAGCGGGCAATCAAGTTGAGAACTTCTTCAAGATTTCCTACATCCGGAT 540
QY 541 GAGAGGTTAGGCTCCCTTCGCCCTTCAGAGCCCAAGCTGGGAGCACTTTCGG 600
Db 541 GAGAGGTTAGGCTCCCTTCGCCCTTCAGAGCCCAAGCTGGGAGCACTTTCGG 600
QY 601 GAGAGGCTCAGCGCCCTTCGGGGCACTTCTCAGGCTCAGAGATTGATTTCCGCTGG 660
Db 601 GAGAGGCTCAGCGCCCTTCGGGGCACTTCTCAGGCTCAGAGATTGATTTCCGCTGG 660
QY 661 ATCCGTGCCCCCACTAGTTTGTGTGTCATCTCGAAGAGAGGGCTTGACATTAAG 720
Db 661 ATCCGTGCCCCCACTAGTTTGTGTGTCATCTCGAAGAGAGGGCTTGACATTAAG 720
QY 721 AACATCACTAAGCAAGCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTTGGAGCT 780
Db 721 AACATCACTAAGCAAGCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTTGGAGCT 780
QY 781 GCGAGAGAGCTGTACCATCATGTCACCCCAAGGGGACTTCTGAAGCATGCCGATG 840
Db 781 GCGAGAGAGCTGTACCATCATGTCACCCCAAGGGGACTTCTGAAGCATGCCGATG 840
QY 841 ATTCTTGAATCATCAGAAAAGGCGAGTGAAGCCAACTAGCCGAAGATTCCTCG 900
Db 841 ATTCTTGAATCATCAGAAAAGGCGAGTGAAGCCAACTAGCCGAAGATTCCTCG 900
QY 901 AAAATCTTGGCAACACATGCTTGTGTGAGACTGATTGGAAAAGAGGCAAAATTG 960
Db 901 AAAATCTTGGCAACACATGCTTGTGTGAGACTGATTGGAAAAGAGGCAAAATTG 960
QY 961 AAGAAAATTGAAACATGAAACAGGACCAAGATPAACATCTCATCTTGGAGATTGAGC 1020
Db 961 AAGAAAATTGAAACATGAAACAGGACCAAGATPAACATCTCATCTTGGAGATTGAGC 1020
QY 1021 AATATCAACCCCGAAGAAACATCACTGTGAAGGGCAAGTTGAGGCTGTGCCAGTGT 1080
Db 1021 AATATCAACCCCGAAGAAACATCACTGTGAAGGGCAAGTTGAGGCTGTGCCAGTGT 1080
QY 1081 GAGATAGAGATTATGAAGAGCTGCTGAGGCTTTGAAAATGATGCTGCTGTTAAC 1140
Db 1081 GAGATAGAGATTATGAAGAGCTGCTGAGGCTTTGAAAATGATGCTGCTGTTAAC 1140
QY 1141 ACCCACTCCGGAACTCTTCCAGCCGTGTAACCCCATCAGAGTTGGCCGTCCGCAT 1200
Db 1141 ACCCACTCCGGAACTCTTCCAGCCGTGTAACCCCATCAGAGTTGGCCGTCCGCAT 1200
QY 1141 ACCCACTCCGGAACTCTTCCAGCCGTGTAACCCCATCAGAGTTGGCCGTCCGCAT 1200
Db 1141 ACCCACTCCGGAACTCTTCCAGCCGTGTAACCCCATCAGAGTTGGCCGTCCGCAT 1200
QY 1201 CATCACTCTTATCAGAGCAGAGATTGTGAATCTTTCATCCCAACCGAGCTGTGGC 1260
Db 1201 CATCACTCTTATCAGAGCAGAGATTGTGAATCTTTCATCCCAACCGAGCTGTGGC 1260
QY 1261 GCCATCATCGGAGAAAGGGGGCAACATCAACAGCTGGCGAGATTCCCGGAGCTCT 1320
Db 1261 GCCATCATCGGAGAAAGGGGGCAACATCAACAGCTGGCGAGATTCCCGGAGCTCT 1320
QY 1321 ATTAAGATTTGCTCCCGAGAAAGGCCAGAGCTCAGAGAAAGATGTGATCATCACCGG 1380
Db 1321 ATTAAGATTTGCTCCCGAGAAAGGCCAGAGCTCAGAGAAAGATGTGATCATCACCGG 1380
QY 1381 CCACCGAAGCCAGATTCAAGGCCCAAGGCGATCTTTGGGAACTGAAAAGAGAAAAC 1440
Db 1381 CCACCGAAGCCAGATTCAAGGCCCAAGGCGATCTTTGGGAACTGAAAAGAGAAAAC 1440
QY 1441 TTCTTTAAACCCCAAGAAAGTGAAGCTGGAAGCGATATCAGAGTCCCTCTTCCACA 1500
Db 1441 TTCTTTAAACCCCAAGAAAGTGAAGCTGGAAGCGATATCAGAGTCCCTCTTCCACA 1500
QY 1501 GCTGCGCGGGTGAATGGCAAGGTGCAAGACGTGAACGAATCGCAAGACTTAAACAGT 1560
Db 1501 GCTGCGCGGGTGAATGGCAAGGTGCAAGACGTGAACGAATCGCAAGACTTAAACAGT 1560
QY 1561 GCAGAACTCATGCTGCTGCTGACCAAAAGCCAGATGAAAATGAGAGTGTGTGAGA 1620
Db 1561 GCAGAACTCATGCTGCTGCTGACCAAAAGCCAGATGAAAATGAGAGTGTGTGAGA 1620

QY 1621 ATTATCGGCACTTCTTGTAGAGCCAGATGCAAGGCAAGATGAGGAAATTGTACAA 1680
Db 1621 ATTATCGGCACTTCTTGTAGAGCCAGATGCAAGGCAAGATGAGGAAATTGTACAA 1680
QY 1681 CAGGTGAAGCAGCAGAGAGCAAAATACCTTCAAGGAGTGTGCTTCAAGGCGAGAGTGA 1740
Db 1681 CAGGTGAAGCAGCAGAGAGCAAAATACCTTCAAGGAGTGTGCTTCAAGGCGAGAGTGA 1740
QY 1741 GGTCCCAACAGGCAACGCAAAACAGGATGAATGTAGCCCTTCAACACTGACAGAA 1800
Db 1741 GGTCCCAACAGGCAACGCAAAACAGGATGAATGTAGCCCTTCAACACTGACAGAA 1800
QY 1801 TGAGACCAAAAGCAGCAGCCAGATCGGAGCAAAACCAAGACATCTGAGAAATGAGAA 1860
Db 1801 TGAGACCAAAAGCAGCAGCCAGATCGGAGCAAAACCAAGACATCTGAGAAATGAGAA 1860
QY 1861 GTCTCGAGAGCGCGCAGGAGCTGTGCGAGAGCTTGAAGAACCCAGGGGCGAGAGAGG 1920
Db 1861 GTCTCGAGAGCGCGCAGGAGCTGTGCGAGAGCTTGAAGAACCCAGGGGCGAGAGAGG 1920
QY 1921 GCGGGAAAGTCAAGCAGTTTGCAGAACCAAGCCGCTCCCGCCCGCAGAGGC 1980
Db 1921 GCGGGAAAGTCAAGCAGTTTGCAGAACCAAGCCGCTCCCGCCCGCAGAGGC 1980
QY 1981 TTCTGCAAGCTTCAAGCCATCCATTCACTGAGATCTTCTGAACTTCCACGA 2040
Db 1981 TTCTGCAAGCTTCAAGCCATCCATTCACTGAGATCTTCTGAACTTCCACGA 2040
QY 2041 CGCTATCCCTTTTATGTTGAATCAATATGAGTGAAGTGTGAAACCAAGCAAAATGAC 2100
Db 2041 CGCTATCCCTTTTATGTTGAATCAATATGAGTGAAGTGTGAAACCAAGCAAAATGAC 2100
QY 2101 ACCCTTTTCTGTGCAATGCTCTGTATCACTGTGTATCATATTAGAAAGGAGATG 2160
Db 2101 ACCCTTTTCTGTGCAATGCTCTGTATCACTGTGTATCATATTAGAAAGGAGATG 2160
QY 2161 TTAAAGATGTGCTGTGAGTTACACAGGAGTGTGCAAGGCTTAATATTTAGAAAT 2220
Db 2161 TTAAAGATGTGCTGTGAGTTACACAGGAGTGTGCAAGGCTTAATATTTAGAAAT 2220
QY 2221 AATATATCAAAATCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 2280
Db 2221 AATATATCAAAATCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 2280
QY 2281 TTAAAGAAAGCAGGCTTTCTAGACTTAAAGATAAGTCTTGGAGGTCAAG 2340
Db 2281 TTAAAGAAAGCAGGCTTTCTAGACTTAAAGATAAGTCTTGGAGGTCAAG 2340
QY 2341 TGTAGAGAGAGCTTGAAGGCAACCGCACAAAATTCACAGAGGAAATCTGTCGGA 2400
Db 2341 TGTAGAGAGAGCTTGAAGGCAACCGCACAAAATTCACAGAGGAAATCTGTCGGA 2400
QY 2401 AGGACACTCAGCGCACTTCTGATCACTGTGTATGTCAACAGAGGATACCGTCTCT 2460
Db 2401 AGGACACTCAGCGCACTTCTGATCACTGTGTATGTCAACAGAGGATACCGTCTCT 2460
QY 2461 TGAAGAGAACTCTGTCACTCTCATGCTGTCTAGCTCATACCCATTTCTCTTGC 2520
Db 2461 TGAAGAGAACTCTGTCACTCTCATGCTGTCTAGCTCATACCCATTTCTCTTGC 2520
QY 2521 TTCAAGGTTTAACTGTTTTTGTGATATGCTATATATTTCTGTCTCTGCT 2580
Db 2521 TTCAAGGTTTAACTGTTTTTGTGATATGCTATATATTTCTGTCTCTGCT 2580
QY 2581 TATCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCAT 2640
Db 2581 TATCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCAT 2640
QY 2641 CCGTCCATCTCAATCCGATCTTACAGCAACCCCGCCCGCCAGGCAAGCAGTCTGGA 2700
Db 2641 CCGTCCATCTCAATCCGATCTTACAGCAACCCCGCCCGCCAGGCAAGCAGTCTGGA 2700

OY		2701	GATATCATTCACACAAAAGGAAACAAAAGGAACAACCAACCAGCTTCAATTGCATT	2767
Db		2701	GTATCATCATCACAAAAGGAAACAAAAGGAACAACCAACCAGCTTCAATTGCATT	2767
OY		2761	GGTTACTCAAAGAACAAGATGCATATGTGATCTTGCCATAGCTTTGGAAAGGAAACA	2820
Db		2761	GGTTACTCAAAGAACAAGATGCATATGTGATCTTGCCATAGCTTTGGAAAGGAAACA	2820
OY		2821	GGAACCCACCAACCAACCAATCAACCAACCAACCAAAAAATTCACATGAAAGATGT	2880
Db		2821	GGAACCCACCAACCAACCAATCAACCAACCAACCAAAAAATTCACATGAAAGATGT	2880
OY		2881	AATTTGCTTTTTGCAATTTTGGTGTATTAAGCATAAATTCAGCAAAATGATTCCTTC	2940
Db		2881	AATTTGCTTTTTGCAATTTTGGTGTATTAAGCATAAATTCAGCAAAATGATTCCTTC	2940
OY		2941	TTTAAAAAATAATGTGAGGAATGTGAATAATTCAGAGGTGGCCACAGGGCTT	3000
Db		2941	TTTAAAAAATAATGTGAGGAATGTGAATAATTCAGAGGTGGCCACAGGGCTT	3000
OY		3001	AAATTCAAGATTTTTTTAACGAGAAAAACACAGAGAAGCTACCTCAGGTGTTTTTA	3060
Db		3001	AAATTCAAGATTTTTTTAACGAGAAAAACACAGAGAAGCTACCTCAGGTGTTTTTA	3060
OY		3061	CCTCAGCACCCTGCTCTGTGTTCCCTTAAGATTTGTAAAGCTGATAGTGAACAT	3120
Db		3061	CCTCAGCACCCTGCTCTGTGTTCCCTTAAGATTTGTAAAGCTGATAGTGAACAT	3120
OY		3121	TTTTTATTTTTTTAAATPAAAAATGAGTGGAAAAAATAAGATATCAACTGCCAGGCTG	3180
Db		3121	TTTTTATTTTTTTAAATPAAAAATGAGTGGAAAAAATAAGATATCAACTGCCAGGCTG	3180
OY		3181	GAGAAAGTGACAGTCCAAGTGTGCACAGCTGTTCGAAATGCTTCGCGTAGCCAGAA	3240
Db		3181	GAGAAAGTGACAGTCCAAGTGTGCACAGCTGTTCGAAATGCTTCGCGTAGCCAGAA	3240
OY		3241	CCNATATGCGCTTCTTTTGGACAAACCTTGAAAAATGTTTATTT	3283
Db		3241	CCNATATGCGCTTCTTTTGGACAAACCTTGAAAAATGTTTATTT	3283
RESULT 3				
Bd209927				
LOCUS		BD209927	3283 bp DNA linear PAT 17-JUL-2003	
DEFINITION			Isolated nucleic acid molecules encoding cancer-associated antigens, these antigens and method of using the same.	
ACCESSION		BD209927		
VERSION		BD209927.1	GI:33019697	
KEYWORDS		JP 2002512049-A/6.		
SOURCE		Homo sapiens (human)		
ORGANISM		Homo sapiens		
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS		Chen,Y.T., Gare,A., Tsang,S., Stockert,E., Jager,E., Knuth,A. and Old,L.J.		
TITLE			Isolated nucleic acid molecules encoding cancer-associated antigens, these antigens and method of using the same	
JOURNAL			Patent: JP 2002512049-A 6-23-APR-2002;	
COMMENT		LUDWIG INSTITUTE FOR CANCER RESEARCH		
		OS Homo sapiens (human)		
		PN JP 2002512049-A/6		
		PD 23-APR-2002		
		PF 16-MAR-1999 JP 2000545030		
		PR 17-APR-1998 US 09/061709		
		PI YAO TSANG CHEN,ALI GURE,SOLAM TSANG,ELISABETH STOCKERT,ELKE PI JAGER,		
		PI ALEXANDER KNUTH,LLOYD J OLD		
		PC C12N15/09,A6K35/12,A6IK39/00,A6IP35/00,C07K16/32,		
		PC C12N1/15,		
		C12N1/19,C12N1/21,C12N5/10,C12P21/08,C12Q1/66,G01N33/53,G01N33/		
		574,		
		PC		

FEATURES		location/Qualifiers	
source		1..3283	/organism="Homo sapiens (human)".
		/organism="Homo sapiens"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:9606"	

ORIGIN

Query Match	Best Local Similarity	100.0%	Score 3282;	DB 6;	Length 3283;
Matches 3283;	Conservative	100.0%	0;	Mismatches	0;
QY	1	GGCAGCGAGGAGGCGAGGAGCGCGGGTACCGGAGCGGGGAGCCGGCGGGCTCTCGGGG	60		
Db	1	GGCAGCGGAGGAGGCGAGGAGCGCGGGTACCGGGCCGGGGAGCCGGGGCTCTCGGGG	60		
QY	61	AAGAGCGGATGATGAACAACTTTTACATCGGGAACCTGAGCCCCCGCGTCAACCGCCGAC	120		
Db	61	AAGAGCGGATGATGAACAACTTTTACATCGGGAACCTGAGCCCCCGCGTCAACCGCCGAC	120		
QY	121	GACCTCGGGAGCGCTTTTGGGGGAGGAGAGTGCCCTCGGCGGAGAGGTCTGCTGAAG	180		
Db	121	GACCTCGGGAGCGCTTTTGGGGGAGGAGAGTGCCCTCGGCGGAGAGGTCTGCTGAAG	180		
QY	181	TCCGGCTACGCGCTTCGTGGACTACCCCGACGAGAACTCGGGCATTCGGCGCATCGAGACC	240		
Db	181	TCCGGCTACGCGCTTCGTGGACTACCCCGACGAGAACTCGGGCATTCGGCGCATCGAGACC	240		
QY	241	CTCTCGGGTAAAGTGAATTGTCATGGGAAAATCATGAGTTGATTACTACGTCTTAA	300		
Db	241	CTCTCGGGTAAAGTGAATTGTCATGGGAAAATCATGAGTTGATTACTACGTCTTAA	300		
QY	301	AAGCTAAGGAGCAGGAAAATTCAATTGCAAACTCCCTCTCACTCGACAGTGGAGAGTG	360		
Db	301	AAGCTAAGGAGCAGGAAAATTCAATTGCAAACTCCCTCTCACTCGACAGTGGAGAGTG	360		
QY	361	TTGATGATGACTTTTGGCTCAATATGAGGACGTGGAGAAATGTGAAACAAGTCAACAGAC	420		
Db	361	TTGATGATGACTTTTGGCTCAATATGAGGACGTGGAGAAATGTGAAACAAGTCAACAGAC	420		
QY	421	ACAGAAACCGCCGTTGTCAACGTCAACATATGCAACAAGAGAAAGCAAAAATAGCCATG	480		
Db	421	ACAGAAACCGCCGTTGTCAACGTCAACATATGCAACAAGAGAAAGCAAAAATAGCCATG	480		
QY	481	GAGAAAGCTAAGCGGGCATCAGTTTGAGAACTACTCTTCAAGATTTCTACATCCGGAT	540		
Db	481	GAGAAAGCTAAGCGGGCATCAGTTTGAGAACTACTCTTCAAGATTTCTACATCCGGAT	540		
QY	541	GAAAGGTGAGCTCCCTTCGCCCCCTCAAGCGAGCCAGGTGGGACCACTCTTCCCGG	600		
Db	541	GAAAGGTGAGCTCCCTTCGCCCCCTCAAGCGAGCCAGGTGGGACCACTCTTCCCGG	600		
QY	601	GAGCAAGGCCACGCCCTCGGGGSCATTTCAAGCGACAGACATTAATTTCCGCTCGCGG	660		
Db	601	GAGCAAGGCCACGCCCTCGGGGSCATTTCAAGCGACAGACATTAATTTCCGCTCGCGG	660		
QY	661	ATCCTGTGCTCCCAACCAATTTTGTGTGCATCATGGAAGAGAGGGCTTGAACATAAAG	720		
Db	661	ATCCTGTGCTCCCAACCAATTTTGTGTGCATCATGGAAGAGAGGGCTTGAACATAAAG	720		
QY	721	AACATCACTAAGAGACCAAGTCCCGGGTATGATATCCATAGAAAAGAACTCTGGAAGCT	780		
Db	721	AACATCACTAAGAGACCAAGTCCCGGGTATGATATCCATAGAAAAGAACTCTGGAAGCT	780		
QY	781	GCAGAAAGCTGTCAACATTCATGACACCCAGAGGGGACCTTGTAAACATAGCCGACAG	840		
Db	781	GCAGAAAGCTGTCAACATTCATGACACCCAGAGGGGACCTTGTAAACATAGCCGACAG	840		
QY	841	ATTCTTAAATCATGCAAGAAAGGCGAGATGAGCAAACTAAGCCGAAAGATTCCTCTG	900		

Db 841 ATTTCTTGAATCATGCGAAGAGCGACATGAGCAAACTAGCGGAAGAAATTCCTCTG 900
Qy 901 AAAATCTTGACACAAATGCTTGTTGGAAGACTGATTTGGAAAAAGAGCAGAAATTTG 960
Db 901 AAAATCTTGACACAAATGCTTGTTGGAAGACTGATTTGGAAAAAGAGCAGAAATTTG 960
Qy 961 AAGAAATTTGAACATGAAACAGGGACCAAGATACATCTCTTTGACAGATTTGAGC 1020
Db 961 AAGAAATTTGAACATGAAACAGGGACCAAGATACATCTCTTTGACAGATTTGAGC 1020
Qy 1021 AATATCAACCCGGAAGAAAGCACTGTGAAGGGCAGAGTTGAGGCTGTGCCAGTGT 1080
Db 1021 AATATCAACCCGGAAGAAAGCACTGTGTGAAGGGCAGAGTTGAGGCTGTGCCAGTGT 1080
Qy 1081 GAGATAGAGATTAATGAAAGAGCTGTGAGGCTTTGAAATGATATGCTGTGTTAAC 1140
Db 1081 GAGATAGAGATTAATGAAAGAGCTGTGAGGCTTTGAAATGATATGCTGTGTTAAC 1140
Qy 1141 ACCCACTCCGATCTTCTCCAGCTGTACCCCATCAACAGTTTGGCCCGGAT 1200
Db 1141 ACCCACTCCGATCTTCTCCAGCTGTACCCCATCAACAGTTTGGCCCGGAT 1200
Qy 1201 CATGCTCTTATCCAGAGAGAGATTTGATCTCTTCATCCCAACCGAGCTGTGGC 1260
Db 1201 CATGCTCTTATCCAGAGAGAGATTTGATCTCTTCATCCCAACCGAGCTGTGGC 1260
Qy 1261 GGCATCATCGGAGAAAGGGGGCACAATCAACAGCTGGAGATTGCGCGAGCTCT 1320
Db 1261 GGCATCATCGGAGAAAGGGGGCACAATCAACAGCTGGAGATTGCGCGAGCTCT 1320
Qy 1321 ATCAAGATTTGCTTCCGGAAGGCCCAAGCTGTCAAGGAAAGATGCTCATCAACCGGG 1380
Db 1321 ATCAAGATTTGCTTCCGGAAGGCCCAAGCTGTCAAGGAAAGATGCTCATCAACCGGG 1380
Qy 1381 GCACCGGAAGCCCAAGTTCAAGGCCCAAGAGCATTTTGGGAAACTGAAAGAGAAAAAC 1440
Db 1381 GCACCGGAAGCCCAAGTTCAAGGCCCAAGAGCATTTTGGGAAACTGAAAGAGAAAAAC 1440
Qy 1441 TTCTTTAAACCCCAAGAAAGTGAAGCTGAAGGCGCATATCAAGTGCCTTTCCACA 1500
Db 1441 TTCTTTAAACCCCAAGAAAGTGAAGCTGAAGGCGCATATCAAGTGCCTTTCCACA 1500
Qy 1501 GCTGCGCGGGGTGATTTGGCAAAAGTGTGCAAGCCGTGAAGAACTGAGAACTTAAACAGT 1560
Db 1501 GCTGCGCGGGGTGATTTGGCAAAAGTGTGCAAGCCGTGAAGAACTGAGAACTTAAACAGT 1560
Qy 1561 GCAGAGTCACTGTGCTGTGTGCAAAAGGCTGCAAGCCGATGAAATGAGGAAGTGTCAAG 1620
Db 1561 GCAGAGTCACTGTGCTGTGTGCAAAAGGCTGCAAGCCGATGAAATGAGGAAGTGTCAAG 1620
Qy 1621 AATTATCGGGCACTCTTGTGTAGCCAGACTGTCAAGCGCAAGTCAAGGAAATTTGACAA 1680
Db 1621 AATTATCGGGCACTCTTGTGTAGCCAGACTGTCAAGCGCAAGTCAAGGAAATTTGACAA 1680
Qy 1681 CAGGTGAAGAGAGAGCAAGAAATACCTCAAGGAGTGCCTCAAGCGCAAGAGTGA 1740
Db 1681 CAGGTGAAGAGAGAGCAAGAAATACCTCAAGGAGTGCCTCAAGCGCAAGAGTGA 1740
Qy 1741 GGTCTCCACAGGACCGCAAAACAAGGATGATGAGCTTCCAAACCTTGACAGAA 1800
Db 1741 GGTCTCCACAGGACCGCAAAACAAGGATGATGAGCTTCCAAACCTTGACAGAA 1800
Qy 1801 TGAAGCAAAACGAGCCAGATCGGAGCAAAACCAAGACCATTTGAGAAATGAGAA 1860
Db 1801 TGAAGCAAAACGAGCCAGATCGGAGCAAAACCAAGACCATTTGAGAAATGAGAA 1860
Qy 1861 GTCTGCGAAGCGCCAGGAGATCTTGTGCGAGGCTCTAAGAACCCCAAGAGGCGAGAGGG 1920
Db 1861 GTCTGCGAAGCGCCAGGAGATCTTGTGCGAGGCTCTAAGAACCCCAAGAGGCGAGAGGG 1920
Qy 1921 GCGGGGAAGGTACGCGAGTTTGTGCGAAGCAACGAGCCCGCTCCCGGCCCCCAAGGC 1980
Db 1921 GCGGGGAAGGTACGCGAGTTTGTGCGAAGCAACGAGCCCGCTCCCGGCCCCCAAGGC 1980

Db 1921 GCGGGGAAGGTACGCGAGTTTGTGCGAAGCAACGAGCCCGCTCCCGGCCCCCAAGGC 1980
Qy 1981 TTTGCGAGGGCTTGAAGCATCCACTTCAACATCACTCGGATCTCTCTGAATCCACGA 2040
Db 1981 TTTGCGAGGGCTTGAAGCATCCACTTCAACATCACTCGGATCTCTCTGAATCCACGA 2040
Qy 2041 CGCTATCCCTTTTGAATGAACTAAATAGTGAAGGTGTCAAAAGCCAGAAATGCAAC 2100
Db 2041 CGCTATCCCTTTTGAATGAACTAAATAGTGAAGGTGTCAAAAGCCAGAAATGCAAC 2100
Qy 2101 ACCCTTTTCTGTGCAAAATGCTCTGTGTACATGTGTGTACATTTAAGAAAGGAAGT 2160
Db 2101 ACCCTTTTCTGTGCAAAATGCTCTGTGTACATGTGTGTACATTTAAGAAAGGAAGT 2160
Qy 2161 TTAAGATATGTGGCCGTGTGAGTTACACAGGCTGTGACGGGTATATATTTTGAAT 2220
Db 2161 TTAAGATATGTGGCCGTGTGAGTTACACAGGCTGTGACGGGTATATATTTTGAAT 2220
Qy 2221 AATATATCAAAATTAACCTCAACTCAATTTTATCAATTTATTTATTTTCTTTT 2280
Db 2221 AATATATCAAAATTAACCTCAACTCAATTTTATCAATTTATTTATTTTCTTTT 2280
Qy 2281 TTAAGAAAGAAAGAGCTTTTCTAGCTTTAAGAAATTAAGCTTTGGAGGTCTCAAG 2340
Db 2281 TTAAGAAAGAAAGAGCTTTTCTAGCTTTAAGAAATTAAGCTTTGGAGGTCTCAAG 2340
Qy 2341 TTAAGAAAGAAAGAGCTTTTCTAGCTTTAAGAAATTAAGCTTTGGAGGTCTCAAG 2400
Db 2341 TTAAGAAAGAAAGAGCTTTTCTAGCTTTAAGAAATTAAGCTTTGGAGGTCTCAAG 2400
Qy 2401 AGGACACTCAAGGCACTTGTGATCACTGTGTATGTCAAGAAAGGATACCGTCTCT 2460
Db 2401 AGGACACTCAAGGCACTTGTGATCACTGTGTATGTCAAGAAAGGATACCGTCTCT 2460
Qy 2461 TGAAGAGAAACTCTGTACCTCTCATGCTGTCTAGCTATACCACTTTCTCTTGC 2520
Db 2461 TGAAGAGAAACTCTGTACCTCTCATGCTGTCTAGCTATACCACTTTCTCTTGC 2520
Qy 2521 TTCAAGGTTTAACTGTTTGTGCTATGCTATATATTTCTGTCTCTCTGTT 2580
Db 2521 TTCAAGGTTTAACTGTTTGTGCTATGCTATATATTTCTGTCTCTCTGTT 2580
Qy 2581 TATCTCTCCCT 2640
Db 2581 TATCTCTCCCT 2640
Qy 2641 CCTCTCATCTCAATCCGATCTAGAGCACCCCCCCCCCAGGCAAGAGTGTCTGA 2700
Db 2641 CCTCTCATCTCAATCCGATCTAGAGCACCCCCCCCCCAGGCAAGAGTGTCTGA 2700
Qy 2701 GATACATCAACAAAGAAAGAAAGGAAACACAAACCAAGCTCAACTTCACTT 2760
Db 2701 GATACATCAACAAAGAAAGAAAGGAAACACAAACCAAGCTCAACTTCACTT 2760
Qy 2761 GGTACTCAAAAGACAAAGTCAATGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 2820
Db 2761 GGTACTCAAAAGACAAAGTCAATGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 2820
Qy 2821 GGAACCCACAAACCAACCAATCAACCAACCAAGAAATTTCCAAATGAAGATGT 2880
Db 2821 GGAACCCACAAACCAACCAATCAACCAACCAAGAAATTTCCAAATGAAGATGT 2880
Qy 2881 AATTGTCTTTTGTGATTTGTGTATTAAGCATCAATATTCAGCAAAATGATCTCTTTC 2940
Db 2881 AATTGTCTTTTGTGATTTGTGTATTAAGCATCAATATTCAGCAAAATGATCTCTTTC 2940
Qy 2941 TTTAAAAAATAATGTGAGGAAAGTGAAGAAATTTACCAAGTGTGTGCGCCAGGGCTT 3000
Db 2941 TTTAAAAAATAATGTGAGGAAAGTGAAGAAATTTACCAAGTGTGTGCGCCAGGGCTT 3000
Qy 3001 AAATTCACAAATTTTAAACGAAAGAAACACAGAGAAAGTACCTCAGGTGTCTTTTA 3060
Db 3001 AAATTCACAAATTTTAAACGAAAGAAACACAGAGAAAGTACCTCAGGTGTCTTTTA 3060

QY 3061 CCTGACACCTGCTCTGTTGTTCCCTAGAGATTTTGTAAAGCTGATGTTGAGCAT 3120
Db 3061 CCTGACACCTGCTCTGTTGTTCCCTAGAGATTTTGTAAAGCTGATGTTGAGCAT 3120
QY 3121 TTTTATTTTATTTTATTAATAAAGTTGAGAAAAAATAAGATATCAATGCCAGCTTG 3180
Db 3121 TTTTATTTTATTTTATTAATAAAGTTGAGAAAAAATAAGATATCAATGCCAGCTTG 3180
QY 3181 GAGAAAGTGACAGTCCAGTGTGCAACAGCTGTTCTGAAATGTTCTCCGCTAGCCAGAA 3240
Db 3181 GAGAAAGTGACAGTCCAGTGTGCAACAGCTGTTCTGAAATGTTCTCCGCTAGCCAGAA 3240
QY 3241 CCNATATGSCCTCTTTTGACAAACCTTGAAATGTTATTT 3283
Db 3241 CCNATATGSCCTCTTTTGACAAACCTTGAAATGTTATTT 3283

RESULT 4 3667 bp mRNA linear PRI 23-MAY-1999
LOCUS AF057352
DEFINITION Homo sapiens hepatocellular carcinoma autoantigen (p62) mRNA,
complete cds.
ACCESSION AF057352
VERSION AF057352.1 GI:4883680
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM

REFERENCE
AUTHORS Zhang, J.Y., Chan, E.K., Peng, X.X. and Tan, E.M.
TITLE A novel cytoplasmic protein with RNA-binding motifs is an autoantigen in human hepatocellular carcinoma
JOURNAL J. Exp. Med. 189 (7), 1101-1110 (1999)
MEDLINE 99207072
PUBMED 10190901
REFERENCE 2 (bases 1 to 3667)
AUTHORS Zhang, J.Y., Chan, E.K.L., Peng, X.X. and Tan, E.M.
TITLE Direct Submission
JOURNAL Submitted (03-APR-1998) Molecular & Experimental Medicine, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037, USA

FEATURES
source location/Qualifiers

gene 1..3667
CDS /organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..3667
/gene="p62"
436..2106
/gene="p62"
/note="RNA-binding protein"
/codon_start=1
/product="hepatocellular carcinoma autoantigen"
/protein_id="AAU31596.1"
/db_xref="GI:4883681"
/translation="MMKKYTGUSPAVYADDIQQLFGDRKFLPAGQVLTKSGYAVD
YPLDMALRIETLSGVEVHAGKIMEVDVSVSKRLKQINIPHLQEWELDGL
AQYGVENEQVETVETAVNTVATREAKTAMEKLSCHQENISFKSYLPDEEV
SSPSPPQARQDGHSSREGHAGCGTSQAOIDPRLIIVPTQFVAGIIEKELITKN
ITKOTOSVDIHRKENSREGHAGCGTSQAOIDPRLIIVPTQFVAGIIEKELITKN
IKTLANGVAGRLIKEGRRLKKIEHRTGKTTISLQDSTYNPERITTVKTVKAC
ASAETIMKLRALAFENDMLAVNTHSGYRSLSLPHHQGFEPHHSYPEQELVNLFI
TQAVGATIGKGAHITQALAFAGASIKIAEAGPDVSEKRVIIITGPEAQFKAQGF
GKIKENFNPKEKVKLEAHIRVPSVAGVIGKGTVEILQNLTAIEVIVRDQTP
DENEEVIVRIIGHFPAQSOTAKRIREIVQVKQOQKVPQGVASQSK"

Query Match 99.1%; Score 3253.8; DB 9; Length 3667;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 3277; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 3 CAGCGAGAGAGAGAGAGAGCGCGCGGTTACCGGCGCGGGGGAGCCCGGGGCTCTCGGGGAA 62
Db 369 CAGCGAGAGAGAGAGAGCGCGCGGTTACCGGCGCGGGGGAGCCCGGGGCTCTCGGGGAA 428
QY 63 GAGACGATGATGAACAAGCTTTTACATCGGGAACTTGAGCCCGCGCTCAACCGCGAGA 122
Db 429 GAGACGATGATGAACAAGCTTTTACATCGGGAACTTGAGCCCGCGCTCAACCGCGAGA 488
QY 123 CTTCCGAGAGCTCTTTGGGGACAGAGAGCTGCCCCCTGGCGGGACAGTCTGCTGAAGTC 182
Db 489 CTTCCGAGAGCTCTTTGGGGACAGAGAGCTGCCCCCTGGCGGGACAGTCTGCTGAAGTC 548
QY 183 CGGCTAGCGCTTCTGAGACTACCCCGACCAAACTGCGGCATTCGCGGCATCGAGACCT 242
Db 549 CGGCTAGCGCTTCTGAGACTACCCCGACCAAACTGCGGCATTCGCGGCATCGAGACCT 608
QY 243 CTGCGGTAAAGTGAATTCATGAGGAAATCATGGAATGATTAAGTCACTGCTGTAATAA 302
Db 609 CTGCGGTAAAGTGAATTCATGAGGAAATCATGGAATGATTAAGTCACTGCTGTAATAA 668
QY 303 GCTAAGAGAGAGAGAAATTCAGATTGAGAAATCCCTGCTGACCTGAGTGGAGGTGT 362
Db 669 GCTAAGAGAGAGAGAAATTCAGATTGAGAAATCCCTGCTGACCTGAGTGGAGGTGT 728
QY 363 GGATGAGCTTTTGGCTCAATATGGGACAGTGGAGAAATGCTGGAACAAGTCAACAGACAC 422
Db 729 GGATGAGCTTTTGGCTCAATATGGGACAGTGGAGAAATGCTGGAACAAGTCAACAGACAC 788
QY 423 AGAAACCGCGCTTGTCAACGTCATATGCAACAAGAGAGAGACAAATATAGCCATGGA 482
Db 789 AGAAACCGCGCTTGTCAACGTCATATGCAACAAGAGAGAGACAAATATAGCCATGGA 848
QY 483 GAAGCTAAGCGGCGATGATTTGAGAACTACTCCTTCAAGATTTCTATCATCCGGATGA 542
Db 849 GAAGCTAAGCGGCGATGATTTGAGAACTACTCCTTCAAGATTTCTATCATCCGGATGA 908
QY 543 AGAGGTAGGTCCTCCCTGCGCCCTCGACGAGCCAGCTGGGGACCACTTCCCGGGA 602
Db 909 AGAGGTAGGTCCTCCCTGCGCCCTCGACGAGCCAGCTGGGGACCACTTCCCGGGA 968
QY 603 GCAAGGCGACGCGCTTGGGGGCACTTCTAGGCGCAGACAGATTGATTTCCCGTGGGAT 662
Db 969 GCAAGGCGACGCGCTTGGGGGCACTTCTAGGCGCAGACAGATTGATTTCCCGTGGGAT 1028
QY 663 CTTGTCCTCCACCCGATGTTGTTGGTGCATCATCGAAGAGGCGCTTGAACATTAAGAA 722
Db 1029 CTTGTCCTCCACCCGATGTTGTTGGTGCATCATCGAAGAGGCGCTTGAACATTAAGAA 1088
QY 723 CATCACTAAGCAGACCGAGTCCCGGGTATGATTCATAGAAAAGAACTTGGAGCTGC 782
Db 1089 CATCACTAAGCAGACCGAGTCCCGGGTATGATTCATAGAAAAGAACTTGGAGCTGC 1148
QY 783 AGAGAGGCGTGTACCATCCATCCACCCAGAGGGGACTTGTGAACATGCGCATGAT 842
Db 1149 AGAGAGGCGTGTACCATCCATCCACCCAGAGGGGACTTGTGAACATGCGCATGAT 1208
QY 843 TCTTGAATCATGCAAGAAAGGCAATGAGACCAATCTGACCGAAGAGATTCTCTGAA 902
Db 1209 TCTTGAATCATGCAAGAAAGGCAATGAGACCAATCTGACCGAAGAGATTCTCTGAA 1268
QY 903 AATCTTGGCAGCAAGAGGCTTGGTGGAGATCGATTGGAAAAGAGGCAAAATTTGAA 962
Db 1269 AATCTTGGCAGCAAGAGGCTTGGTGGAGATCGATTGGAAAAGAGGCAAAATTTGAA 1328
QY 963 GAAATTTGAACATGAACAGAGGACCAAGATTAACATCTCATCTTTGAGAGATTGAGCAT 1022
Db 1329 GAAATTTGAACATGAACAGAGGACCAAGATTAACATCTCATCTTTGAGAGATTGAGCAT 1388
QY 1023 ATACAAACCGGAAAGAACCATCATCTGTGAAGGGCAGAGTTGAGGCTGTGCACTGCTGA 1082
Db 1389 ATACAAACCGGAAAGAACCATCATCTGTGAAGGGCAGAGTTGAGGCTGTGCACTGCTGA 1448
QY 1083 GATAGAGATTATGAAGAGCGTGTGAGGCTTTGAAAATGATATGCTGCTGTAAAC 1142

1449 GATGAGATTATGAGAGAGCTGCGTAGAGCCCTTTGAAAATGATATGCTGGCTTTAAAC 1508
1143 CCACTCCGGATATCTTCTCAGCGCTGTACCCCATCACCAGTTTGGCCGTTCCGCATCA 1202
1509 CCACTCCGGATATCTTCTCAGCGCTGTACCCCATCACCAGTTTGGCCGTTCCGCATCA 1568
1203 TCACCTTATCCAGAGCAGAGATTGTGATCTTCAATCCCAACCCAGGCGTGTGGGCGC 1262
1569 TCACCTTATCCAGAGCAGAGATTGTGATCTTCAATCCCAACCCAGGCGTGTGGGCGC 1628
1263 CATCATCGGAAAGAGAGGCGGCAACATCAAAACAGCTGGCGAGATTGCGCGAGCCTCTAT 1322
1629 CATCATCGGAAAGAGAGGCGGCAACATCAAAACAGCTGGCGAGATTGCGCGAGCCTCTAT 1688
1323 CAAGATTGCCCCCTGGGAAAGCCCAAGCTGCAGCAAAAGATGTCTATCATACCGGGCC 1382
1689 CAAGATTGCCCCCTGGGAAAGCCCAAGCTGCAGCAAAAGATGTCTATCATACCGGGCC 1748
1383 ACCGGAAGCCCGATTCAAGGCCCAAGGACGGATCTTTGGGAAACTGAAAGAGAAAACCT 1442
1749 ACCGGAAGCCCGATTCAAGGCCCAAGGACGGATCTTTGGGAAACTGAAAGAGAAAACCT 1808
1443 CTTTAAACCCAAAGAAAGTGAAGCTGGAGCGCATATCAGAGTGCCCTCTTCCACAGC 1502
1809 CTTTAAACCCAAAGAAAGTGAAGCTGGAGCGCATATCAGAGTGCCCTCTTCCACAGC 1868
1503 TGGCCGCGGTGATTGGCAAAAGTTGGCAAGCCGTGAACGAATCTGCAAGATTGAACAGTGC 1562
1869 TGGCCGCGGTGATTGGCAAAAGTTGGCAAGCCGTGAACGAATCTGCAAGATTGAACAGTGC 1928
1563 ABAAGTCATCTGTCCTCGTGAACCAACCGCAGTGAATGAAGAAATGAGAAATGATGTCAGAA 1622
1929 ABAAGTCATCTGTCCTCGTGAACCAACCGCAGTGAATGAAGAAATGAGAAATGATGTCAGAA 1988
1623 TATCGGSCACTTCTTTGTAGCCAGACTGCAGACGCAAGATCAGGGAATTTGTACACA 1682
1989 TATCGGSCACTTCTTTGTAGCCAGACTGCAGACGCAAGATCAGGGAATTTGTACACA 2048
1683 GGTGAAGCAGCAGAGCAGAAATACCTTACAGGAGTGCCTTCAACGCGCAGCAAGTGAAG 1742
2049 GGTGAAGCAGCAGAGCAGAAATACCTTACAGGAGTGCCTTCAACGCGCAGCAAGTGAAG 2108
1743 CTCGCCAAGGCGCCAGCAAAACAAAGGATGTAATGTATGCCCTTCAACACTGACAGATG 1802
2109 CTCGCCAAGGCGCCAGCAAAACAAAGGATGTAATGTATGCCCTTCAACACTGACAGATG 2168
1803 AGACCAAAAGCAGCCAGCAGATCGGAGCAAAACCAAGCACTGAGGAATGAGAAGT 1862
2169 AGACCAAAAGCAGCCAGCAGATCGGAGCAAAACCAAGCACTGAGGAATGAGAAGT 2228
1863 CTCGGAAGGCGCCAGGAGACTCTGCGAGGCCCTTGAAACCCCAAGGGCGCGAGAGGGG 1922
2229 CTCGGAAGGCGCCAGGAGACTCTGCGAGGCCCTTGAAACCCCAAGGGCGCGAGAGGGG 2288
1923 GGGGAAAGTTCAGCCAGGTTTGGCAAGAACCAAGAGCCCGCTCCGCGCCCAAGGGCTT 1982
2289 GGGGAAAGTTCAGCCAGGTTTGGCAAGAACCAAGAGCCCGCTCCGCGCCCAAGGGCTT 2348
1983 CTCGAGGCTTCAAGCATTCACTTCAACCATCACTCGGATCTCTGAACTCCCAAGAG 2042
2349 CTCGAGGCTTCAAGCATTCACTTCAACCATCACTCGGATCTCTGAACTCCCAAGAG 2408
2043 CTATCCCTTTAATGTAATCAATAGTGAAGTGTTCAAAGCCCAAGCAAAATGACAC 2102
2409 CTATCCCTTTAATGTAATCAATAGTGAAGTGTTCAAAGCCCAAGCAAAATGACAC 2468
2103 CCTTTTCTGAGCAATGCTCTCTGTACATGTGTACATATTAGAAAGGAAGATGT 2162
2469 CCTTTTCTGAGCAATGCTCTCTGTACATGTGTACATATTAGAAAGGAAGATGT 2528
2163 AAGATATGTGGCCGTGGGTTTACACAGGCTGCTGCAGCGGTATATATTTTGAAGAA 2222

2529 AAGATATGTGGCCGTGGGTTTACACAGGCTGCTGCAGCGGTATATATTTTGAAGAA 2588
2223 TATATCAAAATACCAACTCACTCAATTTTATCAATTAATTAATTTTCTTTT 2282
2589 TATATCAAAATACCAACTCACTCAATTTTATCAATTAATTAATTTTCTTTT 2648
2283 AAAAGAAAGCAGGCTTTTCTAGAATTTAAAGAAATTAAGTCTTTGGAGGTCTCAGGTTG 2342
2649 AAAAGAAAGCAGGCTTTTCTAGAATTTAAAGAAATTAAGTCTTTGGAGGTCTCAGGTTG 2708
2343 TAGAGAGAGCTTTGAGGCCACCCGCAAAAATTCACCAAGGAAATCTCGTGGAG 2402
2709 TAGAGAGAGCTTTGAGGCCACCCGCAAAAATTCACCAAGGAAATCTCGTGGAG 2768
2403 GACACTCAGGCAAGTTCTGSAATCACTGTGTATGTCAACAGAGGATTAACGCTCTCTTG 2462
2769 GACACTCAGGCAAGTTCTGSAATCACTGTGTATGTCAACAGAGGATTAACGCTCTCTTG 2828
2463 AAGAGAAACCTCTGACTCTGATGCTGTACTGACTATACCCATTTCTTTGGCTT 2522
2829 AAGAGAAACCTCTGACTCTGATGCTGTACTGACTATACCCATTTCTTTGGCTT 2888
2523 CACAGTTTAACTGTTTTTTTGAATCTGCTATATATATTTCTGCTCTCTGTTA 2582
2889 CACAGTTTAACTGTTTTTTTGAATCTGCTATATATATTTCTGCTCTCTGTTA 2948
2583 TCTCTCCCTCCCTCCCTCCCTCTTCTCCATCTCCATCTTTTGAATTTCTGTATCC 2642
2949 TCTCTCCCTCCCTCCCTCCCTCTTCTCCATCTCCATCTTTTGAATTTCTGTATCC 3008
2643 CTCATCTCATCTCGGTATCTTAGCAACCCCGCCCGCCAGGCAAAAGAGTCTGAA 2702
3009 CTCATCTCATCTCGGTATCTTAGCAACCCCGCCCGCCAGGCAAAAGAGTCTGAA 3067
2703 ATCAATCAACAAGAAAGAAAGCAAGCAACAAACAGCTCAACTTACACTTGA 2762
3069 ATCAATCAACAAGAAAGAAAGCAAGCAACAAACAGCTCAACTTACACTTGA 3127
2763 TTAATCAAAAGAAAGAAAGTCAATGTATCTTGTCTGAGGTTTGGAGAGAAACAG 2822
3128 TTAATCAAAAGAAAGAAAGTCAATGTATCTTGTCTGAGGTTTGGAGAGAAACAG 3187
2823 AACCCAAACCAACCAATCAACCAACCAAGAAAGAAATTTCCAAATGAAGATGAT 2882
3188 AACCCAAACCAACCAATCAACCAACCAAGAAAGAAATTTCCAAATGAAGATGAT 3247
2883 TTTGCTTTTTCATTTGATTTGATATGATCAATATTCAGCAAAATGATTCCTTCTT 2942
3248 TTTGCTTTTTCATTTGATTTGATATGATCAATATTCAGCAAAATGATTCCTTCTT 3307
2943 T-AAAAAAGAAAGTGAAGAAAGTGAATTTTCAAGGTTTGGCCAGAGCGTTA 3001
3308 TAAAAAAGAAAGTGAAGAAAGTGAATTTTCAAGGTTTGGCCAGAGCGTTA 3367
3002 AATTCAAGATTTTATTAACGAGAAACCAACAGAGAGTAACTCAGGTGTTTATC 3061
3368 AATTCAAGATTTTATTAACGAGAAACCAACAGAGAGTAACTCAGGTGTTTATC 3427
3062 CTCAGACCTGCTCTGTTGTTCCCTTAGAGATTTTGAAGCTGATAGTTGAGACAT 3121
3428 CTCAGACCTGCTCTGTTGTTCCCTTAGAGATTTTGAAGCTGATAGTTGAGACAT 3487
3122 TTTTATTTTTTAAATAAATGATGGAAGAAATTAAGATATCACTGACGCTGG 3181
3488 TTTTATTTTTTAAATAAATGATGGAAGAAATTAAGATATCACTGACGCTGG 3547
3182 AGAAGGTGACAGTCCAGGTGCAACAGCTGTTCTGAATGCTTCGCGTAGGCAAGAC 3241
3548 AGAAGGTGACAGTCCAGGTGCAACAGCTGTTCTGAATGCTTCGCGTAGGCAAGAC 3607
3242 CNAATATGACCTCTTTTGAACAACTGAAATGTTATTT 3283
3608 CTATATGACCTCTTTTGAACAACTGAAATGTTATTT 3649

RESULT 5
ARI171865 3412 bp DNA linear PAT 17-DEC-2001
LOCUS Sequence 6 from patent US 6297364.
ACCESSION ARI171865
VERSION ARI171865.1 GI:17910815
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 3412)
Chen, Y.-T., Gure, A., Tsang, S., Stockert, E., Jager, E., Alexander, K.
and Old, L. J.
Isolated nucleic acid molecule encoding cancer associated antigen,
the antigen itself, and uses thereof
Patent: US 6297364-A 6 02-OCT-2001;
Location/Qualifiers
1..3412
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 95.7%; Score 3143; DB 6; Length 3412;
Best Local Similarity 96.2%; Pred. No. 0;
Matches 3283; Conservative 0; Mismatches 0; Indels 129; Gaps 1;

QY 1 GGCGACGGAGAGCGAGAGCGCGGCTACCGGGCGGGGAGCCGCGGCTCTCGGGG 60
Db 1 GGCGACGGAGAGCGAGAGCGCGGCTACCGGGCGGGGAGCCGCGGCTCTCGGGG 60
QY 61 AAGAGCGGATGATGATCAAGCTTTATCATGCGGGAACCTGAGCCCGCGCTACCGCGGAC 120
Db 61 AAGAGCGGATGATGATCAAGCTTTATCATGCGGGAACCTGAGCCCGCGCTACCGCGGAC 120
QY 121 GACCTCGGCGAGCTCTTTGGGGAGAGAGAGCTGCCCCCTGCGGGACAAGTCTGTGAAG 180
Db 121 GACCTCGGCGAGCTCTTTGGGGAGAGAGAGCTGCCCCCTGCGGGACAAGTCTGTGAAG 180
QY 121 GACCTCGGCGAGCTCTTTGGGGAGAGAGAGCTGCCCCCTGCGGGACAAGTCTGTGAAG 180
Db 121 GACCTCGGCGAGCTCTTTGGGGAGAGAGAGCTGCCCCCTGCGGGACAAGTCTGTGAAG 180
QY 181 TCCGGCTAACGCTTCTGCTGAGTACCCCGACCAAGATGCGGCGATCCGCGCATGAGACC 240
Db 181 TCCGGCTAACGCTTCTGCTGAGTACCCCGACCAAGATGCGGCGATCCGCGCATGAGACC 240
QY 241 CTCCTGGGTAAGAGTAATGATGAGGAAATTCATGAAAGTGAATTAATCTAGTCTTAAA 300
Db 241 CTCCTGGGTAAGAGTAATGATGAGGAAATTCATGAAAGTGAATTAATCTAGTCTTAAA 300
QY 241 CTCCTGGGTAAGAGTAATGATGAGGAAATTCATGAAAGTGAATTAATCTAGTCTTAAA 300
Db 241 CTCCTGGGTAAGAGTAATGATGAGGAAATTCATGAAAGTGAATTAATCTAGTCTTAAA 300
QY 301 AAGCTAAGAGCAAGAAATTCAGATTGAAACATCCCTCCTCACTGAGTGGAGAGTG 360
Db 301 AAGCTAAGAGCAAGAAATTCAGATTGAAACATCCCTCCTCACTGAGTGGAGAGTG 360
QY 361 TTGGATGAGCTTTTGGCTCAATATGAGAGTGGAGATGTGAAACAAGTCAACAGAC 420
Db 361 TTGGATGAGCTTTTGGCTCAATATGAGAGTGGAGATGTGAAACAAGTCAACAGAC 420
QY 421 ACAGAAACCGCCGCTTCAACGTCAATATGCAACAAGAAAGAAATTAAGCATG 480
Db 421 ACAGAAACCGCCGCTTCAACGTCAATATGCAACAAGAAAGAAATTAAGCATG 480
QY 481 GAGAGCTAAGCGGAGCATCAGTTTGAAGACTACTCTTCAAGATTTCTTCAATCCGGAT 540
Db 481 GAGAGCTAAGCGGAGCATCAGTTTGAAGACTACTCTTCAAGATTTCTTCAATCCGGAT 540
QY 541 GAAAGAGTGAAGTCTTCCCTTCCGCGCTCAAGAGCCAGCGTGGGAGCACTCTTCCCGG 600
Db 541 GAAAGAGTGAAGTCTTCCCTTCCGCGCTCAAGAGCCAGCGTGGGAGCACTCTTCCCGG 600
QY 601 GAGCAAGGCGACGCGCTTGGGGGAGCTTCTCAGGCGCAGAGATTTTCCGCTGCGG 660
Db 601 GAGCAAGGCGACGCGCTTGGGGGAGCTTCTCAGGCGCAGAGATTTTCCGCTGCGG 660
QY 661 ATCTGTGCTCCCAACCAAGTTTGTGTGCTCATCGGAAAGAGGGCTTGACCATTAAG 720

Db 661 ATCTGTGCTCCCAACCAAGTTTGTGTGCTCATCGGAAAGAGGGCTTGACCATTAAG 720
QY 721 AACATCACTAAGCAGACCCAGTCCCGGTAGATATTCATAGAAAAGAGAACTCTGAGCT 780
Db 721 AACATCACTAAGCAGACCCAGTCCCGGTAGATATTCATAGAAAAGAGAACTCTGAGCT 780
QY 781 GCAAGAGAGCTGTGACCATCATGATGACCCCAAGAGGGAGCTTCTGAAAGATCCGCATG 840
Db 781 GCAAGAGAGCTGTGACCATCATGATGACCCCAAGAGGGAGCTTCTGAAAGATCCGCATG 840
QY 841 ATTCTTGAATCATGACAGAAAGGAGCATGAGACCAACTAGCCGAAGAGATTCCTCG 900
Db 841 ATTCTTGAATCATGACAGAAAGGAGCATGAGACCAACTAGCCGAAGAGATTCCTCG 900
QY 901 AAAATCTTGGACACATGATGCTGTGTTGGAAGACTGATTTGAAAAGAGAGCAAAATTTG 960
Db 901 AAAATCTTGGACACATGATGCTGTGTTGGAAGACTGATTTGAAAAGAGAGCAAAATTTG 960
QY 961 AAGAAAATTTGAACATGAAACAGGGACCAAGATTAACATCTATCTTTCAGAGATTGAGC 1020
Db 961 AAGAAAATTTGAACATGAAACAGGGACCAAGATTAACATCTATCTTTCAGAGATTGAGC 1020
QY 1021 ATATACAAACCGGAAAGAACCATCATGCTGTAAGGGACAGCTTGAAGGCTGAGAGTCT 1080
Db 1021 ATATACAAACCGGAAAGAACCATCATGCTGTAAGGGACAGCTTGAAGGCTGAGAGTCT 1080
QY 1081 GAGATAGAGATTATGAAAGAGCTGAGAGCTTTGAAATGATATGATGCTGCTTTAAC 1140
Db 1081 GAGATAGAGATTATGAAAGAGCTGAGAGCTTTGAAATGATATGATGCTGCTTTAAC 1140
QY 1141 ----- 1140
Db 1141 CAACAGCAATCTGATCTCCAGGGTTGAACCTCAGCGCATTTGGCATTTTCAACAGGA 1200
QY 1201 CTGTCCGTGCTATCTTCCACAGAGGGCCCCCGAGAGTCCCCCGCTCCCTACAC 1260
Db 1201 CTGTCCGTGCTATCTTCCACAGAGGGCCCCCGAGAGTCCCCCGCTCCCTACAC 1260
QY 1261 -----ACCACTCCGATACTTCTTCCAGCTGTACCCCATCAACAGTTTGGCCG 1191
Db 1261 CCTTCACTACCACTCCGATTACTTCTCCAGCTGTACCCCATCAACAGTTTGGCCG 1191
QY 1192 TTCCCGCATCATCTCTTTATCAGAGAGAGATTTGTAATCTTCTATCCCAACCGAG 1251
Db 1192 TTCCCGCATCATCTCTTTATCAGAGAGAGATTTGTAATCTTCTATCCCAACCGAG 1251
QY 1252 GCTGTGGGCGCATCATCGGGAAGAGGGGCAACATCAACAGCTGGGCGAGATTGGCC 1311
Db 1252 GCTGTGGGCGCATCATCGGGAAGAGGGGCAACATCAACAGCTGGGCGAGATTGGCC 1311
QY 1381 GCTGTGGGCGCATCATCGGGAAGAGGGGCAACATCAACAGCTGGGCGAGATTGGCC 1440
Db 1381 GCTGTGGGCGCATCATCGGGAAGAGGGGCAACATCAACAGCTGGGCGAGATTGGCC 1440
QY 1441 GGAAGCTCTATCAAGATTGCCCCCTGGGGAAGGCCAAGCTCAGCGAAAGATGCTATC 1500
Db 1441 GGAAGCTCTATCAAGATTGCCCCCTGGGGAAGGCCAAGCTCAGCGAAAGATGCTATC 1500
QY 1501 ATACCGGGCACCGGAAGGCCAGTTCAAGGCCAGGAGCGATCTTTGGGAAACTGAAA 1560
Db 1501 ATACCGGGCACCGGAAGGCCAGTTCAAGGCCAGGAGCGATCTTTGGGAAACTGAAA 1560
QY 1561 GAGGAAATCTTCTTAAACCCCAAGAAAGATGAGTGAAGGCGCATATCAGAGTCC 1620
Db 1561 GAGGAAATCTTCTTAAACCCCAAGAAAGATGAGTGAAGGCGCATATCAGAGTCC 1620
QY 1621 TCTTCAACAGCTGGCGGGTGAATTGGCAAAAGTGGCAAGACCGTGAACGAAC 1680
Db 1621 TCTTCAACAGCTGGCGGGTGAATTGGCAAAAGTGGCAAGACCGTGAACGAAC 1680
QY 1681 TTAACCAAGTCAAGATCTGCTGCTGTAACCAAGCGCAAGATTAAGAAAGT 1740
Db 1681 TTAACCAAGTCAAGATCTGCTGCTGTAACCAAGCGCAAGATTAAGAAAGT 1740
QY 1740 ATCTGAGATTTATGAGGAGCTTCTTGTAGCGCAAGTCAAGGCAAGATCAAGGAA 1671
Db 1740 ATCTGAGATTTATGAGGAGCTTCTTGTAGCGCAAGTCAAGGCAAGATCAAGGAA 1671

Db 121 GACCTCCGGCAGCTCTTTGGGGAGAGAACTGCCCTTGCGGAGACAGTCTCTGCTAAG 180
Qy 181 TCCGGCTACGCTTGCTGAGCTAACCCCGACGAACTGGGCCATCCGCGCATCGAGACC 240
Db 181 TCCGGCTACGCTTGCTGAGCTAACCCCGACGAACTGGGCCATCCGCGCATCGAGACC 240
Qy 241 CTCTCGGGGTAAAGTGAATTCATGGAATTCATGGAATTCATGGAATTCATGGAATTC 300
Db 241 CTCTCGGGGTAAAGTGAATTCATGGAATTCATGGAATTCATGGAATTCATGGAATTC 300
Qy 301 AAGCTAAGAGAGAGAAATTCATGGAATTCATGGAATTCATGGAATTCATGGAATTC 360
Db 301 AAGCTAAGAGAGAGAAATTCATGGAATTCATGGAATTCATGGAATTCATGGAATTC 360
Qy 361 TTGATGAGCTTTTGGCTCAATATGGAACATGGAATTCATGGAATTCATGGAATTC 420
Db 361 TTGATGAGCTTTTGGCTCAATATGGAACATGGAATTCATGGAATTCATGGAATTC 420
Qy 421 ACAGAAACCGCGCTTGTCAACGTCAATATGCAACAGAGAAAGAAAGAAATTCAGCAT 480
Db 421 ACAGAAACCGCGCTTGTGTCAACGTCAATATGCAACAGAGAAAGAAATTCAGCAT 480
Qy 481 GAGAAAGCTAAGCGGCGCATCACTTTGAGAACTAATCTTCAAGATTTCTTACATCCGAT 540
Db 481 GAGAAAGCTAAGCGGCGCATCACTTTGAGAACTAATCTTCAAGATTTCTTACATCCGAT 540
Qy 541 GAAAGAGTGAAGCTCCCTTGGCCCCCTTACGAGAGCCGAGCGTGGGAGCACTTTCCGG 600
Db 541 GAAAGAGTGAAGCTCCCTTGGCCCCCTTACGAGAGCCGAGCGTGGGAGCACTTTCCGG 600
Qy 601 GAGCAAGGCGCAGCGCCCTTGGGGGCACTTCTAGGCGCAGACAGATTTGATTTCCGCTGG 660
Db 601 GAGCAAGGCGCAGCGCCCTTGGGGGCACTTCTAGGCGCAGACAGATTTGATTTCCGCTGG 660
Qy 661 ATCTGTGTCCTCCACCGTTTGTGTGTCATCAATCGAAAGAGAGGCTTGACATTAAG 720
Db 661 ATCTGTGTCCTCCACCGTTTGTGTGTCATCAATCGAAAGAGAGGCTTGACATTAAG 720
Qy 721 AACATCTAAGCAGACCCAGTCCCCGGGTAAATTCATGAAAGAAAGAACTCTGAGCT 780
Db 721 AACATCTAAGCAGACCCAGTCCCCGGGTAAATTCATGAAAGAAAGAACTCTGAGCT 780
Qy 781 GCAGAGAGCGCTGACCATCATGCGCACCCAGAGGGGACTTGTGAAGATTCGCGATG 840
Db 781 GCAGAGAGCGCTGACCATCATGCGCACCCAGAGGGGACTTGTGAAGATTCGCGATG 840
Qy 841 ATTCTGAAATTCATGAGAAAGAGCAGATGAGCACTAAGCCGAGAGATTCCTGTG 900
Db 841 ATTCTGAAATTCATGAGAAAGAGCAGATGAGCACTAAGCCGAGAGATTCCTGTG 900
Qy 901 AAAATCTTGGCAGCAATGCTGTGTGGAAGCTGATGGAATTCGAAAGAAAGCAAAATTTG 960
Db 901 AAAATCTTGGCAGCAATGCTGTGTGGAAGCTGATGGAATTCGAAAGAAAGCAAAATTTG 960
Qy 961 AAGAAATTTGAATGAAACAGGGACCAAGATTAACAATCTCATCTTTGAGAGATTTGAGC 1020
Db 961 AAGAAATTTGAATGAAACAGGGACCAAGATTAACAATCTCATCTTTGAGAGATTTGAGC 1020
Qy 1021 ATATCAACCCCGAAGAAACATCACTGTGAAGGGCAGATTTGAGGCTGTGCAAGTCT 1080
Db 1021 ATATCAACCCCGAAGAAACATCACTGTGAAGGGCAGATTTGAGGCTGTGCAAGTCT 1080
Qy 1081 GAGATGAGATTTATGAAGAAAGCTGCGTGGAGGCTTTGAAATATATGCTGCTGTTAAC 1140
Db 1081 GAGATGAGATTTATGAAGAAAGCTGCGTGGAGGCTTTGAAATATATGCTGCTGTTAAC 1140
Qy 1141 ----- 1140
Db 1141 CAACAAGCAATCTGATCCAGGGTTGAACCTCAGCGCATCTTTGCAACAGGA 1200
Qy 1141 ----- 1140

Db 1201 CTGTCCGTATCTCCACAGAGAGGCCCGCGAGCTCCCGCGCTGCCCTTAACAC 1260
Qy 1141 -----ACCACTCCGAAATCTTCTCCAGCTGTACCCCATCACCAGTTTGCCCG 1191
Db 1261 CCTTCACTACCACTCCGGAATCTTCTCCAGCTGTACCCCATCACCAGTTTGCCCG 1320
Qy 1192 TTCCCGGATCAATCACTCTTATCCAGAGCAGAGATTTGAAATCTTCAATCCCAACCG 1251
Db 1321 TTCCCGGATCAATCACTCTTATCCAGAGCAGAGATTTGAAATCTTCAATCCCAACCG 1380
Qy 1252 GCTGTGGCGCATCATGAGGAAGAGGGGACATCAACAAACAGTGGCGGATTTGCGC 1311
Db 1381 GCTGTGGCGCATCATGAGGAAGAGGGGACATCAACAAACAGTGGCGGATTTGCGC 1440
Qy 1312 GAGGCTCTATCAAGATTTGCCCTGCGGAAGCCGAGAGCTGACGAAAGAGATGCTCATC 1371
Db 1441 GAGGCTCTATCAAGATTTGCCCTGCGGAAGCCGAGAGCTGACGAAAGAGATGCTCATC 1500
Qy 1372 ATCAACCGGCGCACCGGAAGCCCACTTCAAGGCCGAGAGCGATCTTTGGAATCTGAAA 1431
Db 1501 ATCAACCGGCGCACCGGAAGCCCACTTCAAGGCCGAGAGCGATCTTTGGAATCTGAAA 1560
Qy 1432 GAGGAAATCTTCTTAACCCCAAGAAAGTGAAGCTGGAAGGCGATATCAAGTGGCC 1491
Db 1561 GAGGAAATCTTCTTAACCCCAAGAAAGTGAAGCTGGAAGGCGATATCAAGTGGCC 1620
Qy 1492 TCTTCAACGCTGGCGGCTGATTTGCAAAAGTGGCAAGACCTGTAAAGCACTGCAAGAC 1551
Db 1621 TCTTCAACGCTGGCGGCTGATTTGCAAAAGTGGCAAGACCTGTAAAGCACTGCAAGAC 1680
Qy 1552 TTAAACAGTGCAGAAATCATGTGCTGTGACCAAAAGCCGATGAAATGAGAAAGTG 1611
Db 1681 TTAAACAGTGCAGAAATCATGTGCTGTGACCAAAAGCCGATGAAATGAGAAAGTG 1740
Qy 1612 ATGTCAGAAATTAACGGGCACTTCTTGTAGCCAGCTGACAGGCCCAAGATCAGGAAA 1671
Db 1741 ATGTCAGAAATTAACGGGCACTTCTTGTAGCCAGCTGACAGGCCCAAGATCAGGAAA 1800
Qy 1672 ATTGTACAACAGGTGAAGCAGCAGAGCAGAAATACCTCAGGAGTGCCTCAGACGC 1731
Db 1801 ATTGTACAACAGGTGAAGCAGCAGAGCAGAAATACCTCAGGAGTGCCTCAGACGC 1860
Qy 1732 AGCAAGTGAAGCTTCCACAGGCAACAGCAAAACAGCGATGATGACCTTCAACAC 1791
Db 1861 AGCAAGTGAAGCTTCCACAGGCAACAGCAAAACAGCGATGATGACCTTCAACAC 1920
Qy 1792 CTGACAGAAATGAAGCCAAAGCAGCCAGCCAGATTCGGAAGCAAAAGCAACCACTGAG 1851
Db 1921 CTGACAGAAATGAAGCCAAAGCAGCCAGCCAGATTCGGAAGCAAAAGCAACCACTGAG 1980
Qy 1852 GAATGAGAAGTCTGCGGAGCGGCAAGGACTGTGCGAGAGCCCTGAGAACCCCAAGGGC 1911
Db 1981 GAATGAGAAGTCTGCGGAGCGGCAAGGACTGTGCGAGAGCCCTGAGAACCCCAAGGGC 2040
Qy 1912 CGAGAGGGGCGGGGAAGGTCAAGCAGGTTTGCAAGCAACCGAGCCCGCTCCGCC 1971
Db 2041 CGAGAGGGGCGGGGAAGGTCAAGCAGGTTTGCAAGCAACCGAGCCCGCTCCGCC 2100
Qy 1972 CCCGAGGGCTTTCGAGGGCTTCAGCCATCCACTTCACTGAGGATCTTCCCTGAA 2031
Db 2101 CCCGAGGGCTTTCGAGGGCTTCAGCCATCCACTTCACTGAGGATCTTCCCTGAA 2160
Qy 2032 CTCGACGAGCGCTATCCCTTTTATGTAATTAAGTGAAGCTGTGTAAGCCAGC 2091
Db 2161 CTCGACGAGCGCTATCCCTTTTATGTAATTAAGTGAAGCTGTGTAAGCCAGC 2220
Qy 2092 AAAATGCAACCTTTTCTGTGCGAAATGTCTCTGTACATGTGTACATATTAAGAA 2151
Db 2221 AAAATGCAACCTTTTCTGTGCGAAATGTCTCTGTACATGTGTACATATTAAGAA 2280
Qy 2152 GGGAAATGTTAAGATATGAGCTGTGGGTTTACAGGGGTGCTGACGGGTAATATAT 2211
Db 2281 GGGAAATGTTAAGATATGAGCTGTGGGTTTACAGGGGTGCTGACGGGTAATATAT 2340

QY 361 TTGATGACTTTTGGCTCAATATGAGACAGTGAATGTGAAACAAGTCAACACAGAC 420
DB 361 TTGATGACTTTTGGCTCAATATGAGACAGTGAATGTGAAACAAGTCAACACAGAC 420
QY 421 AAGAAACCGCGCTGTGTCACTCATATGCAAGAGAAAGAAACAAATATGACATG 480
DB 421 AAGAAACCGCGCTGTGTCACTCATATGCAAGAGAAAGAAACAAATATGACATG 480
QY 481 GAGAGCTAAGCGGCGATCAGTTTGAACCTAATCTCTTCAAGATTTTCTCAATCCCGAT 540
DB 481 GAGAGCTAAGCGGCGATCAGTTTGAACCTAATCTCTTCAAGATTTTCTCAATCCCGAT 540
QY 541 GAAAGGTGAGCTCCCTTCGCGCCCTCAGCGAGCCGAGCGGAGACCACTCTCCGCG 600
DB 541 GAAAGGTGAGCTCCCTTCGCGCCCTCAGCGAGCCGAGCGGAGACCACTCTCCGCG 600
QY 601 GAGCAGGCGCAGCGCCCTTGCGGCGCACTTCTCAGCGCAGACAGATTTTCCCGCTGCG 660
DB 601 GAGCAGGCGCAGCGCCCTTGCGGCGCACTTCTCAGCGCAGACAGATTTTCCCGCTGCG 660
QY 661 ATCTGCTGCCACCCAGTTTGTGTGTCATCATCGGAAAGAGAGGCTTGAACCTAAG 720
DB 661 ATCTGCTGCCACCCAGTTTGTGTGTCATCATCGGAAAGAGAGGCTTGAACCTAAG 720
QY 721 AACATCACTAAGCAGACCCAGTCCCGGAGTATTCATAGAAAGAAAGAACTTGAAGCT 780
DB 721 AACATCACTAAGCAGACCCAGTCCCGGAGTATTCATAGAAAGAAAGAACTTGAAGCT 780
QY 781 GCAGAGAAAGCTGTGACCATCATATGCCACCCGAGAGGAGCTTGAAGATGCCGATG 840
DB 781 GCAGAGAAAGCTGTGACCATCATATGCCACCCGAGAGGAGCTTGAAGATGCCGATG 840
QY 841 ATTCTGAAATCATGCGAAGAGCGAGATGAGACCAACTAGCCGAAAGAAATTCCTGTG 900
DB 841 ATTCTGAAATCATGCGAAGAGCGAGATGAGACCAACTAGCCGAAAGAAATTCCTGTG 900
QY 901 AAAATCTTGGCACAACAATGCTGTGTAAGACTGATTTGAAAAAGAGCAGAAATTTG 960
DB 901 AAAATCTTGGCACAACAATGCTGTGTAAGACTGATTTGAAAAAGAGCAGAAATTTG 960
QY 961 AAGAAATTTGAACATGAAACAGGACCAAGATTAACAATCTCATCTTTGACAGATTTGAGC 1020
DB 961 AAGAAATTTGAACATGAAACAGGACCAAGATTAACAATCTCATCTTTGACAGATTTGAGC 1020
QY 1021 ATATACAAACCCGAGAAAGAACATCATCTGTGAAGGCGCACTTGAAGGCTGTGCGAGTGT 1080
DB 1021 ATATACAAACCCGAGAAAGAACATCATCTGTGAAGGCGCACTTGAAGGCTGTGCGAGTGT 1080
QY 1081 GAGATGAGATTTATGAAGAGCTGCTGAGGCTTTGAAAAATGATATGCTGTGTAAC 1140
DB 1081 GAGATGAGATTTATGAAGAGCTGCTGAGGCTTTGAAAAATGATATGCTGTGTAAC 1140
QY 1141 ----- 1140
DB 1141 CAAACAAGCAATCTGATCCAGGGTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGA 1200
QY 1141 ----- 1140
DB 1201 CTGTCCGTGCTATTCACACAGAGGCCCCGCGAGCTCCCCCGCTGCCCTACAC 1260
QY 1141 -----ACCACTCCGAGTAATCTCTCAGCTGTGAACCCCACTACCAAGTTTGGCCG 1191
DB 1261 CCGTTCACTACCACTCCGAGTAATCTCTCAGCTGTGAACCCCACTACCAAGTTTGGCCG 1320
QY 1192 TTTCCCGATATCATCTTTATCCAGAGCAGAGATTTGAAATCTCTTCATCCCAACCCAG 1251
DB 1321 TTTCCCGATATCATCTTTATCCAGAGCAGAGATTTGAAATCTCTTCATCCCAACCCAG 1380
QY 1252 GCTGTGGGCGCCATCATCGGAGAGAGAGGCGGCACTATCAACAGCTGGGAGATTTGCC 1311
DB 1381 GCTGTGGGCGCCATCATCGGAGAGAGAGGCGGCACTATCAACAGCTGGGAGATTTGCC 1440
QY 1312 GAGGCTCTATCAAGATTTGCCCTTGCGAGAGGCTCAGCGCAAGAAAGATGTGATG 1371
DB 1441 GAGGCTCTATCAAGATTTGCCCTTGCGAGAGGCTCAGCGCAAGAAAGATGTGATG 1500
QY 1372 ATCAACGGGCGCACCGGAAGCCAGTTCAAGGCCAGGACGAGATCTTTGGGAACTGAAA 1431
DB 1501 ATCAACGGGCGCACCGGAAGCCAGTTCAAGGCCAGGACGAGATCTTTGGGAACTGAAA 1560
QY 1432 GAGGAAACTTCTTTAAACCCCAAGAGAGTGAAGCTGGAAGCCATATCAGAGTCCC 1491
DB 1561 GAGGAAACTTCTTTAAACCCCAAGAGAGTGAAGCTGGAAGCCATATCAGAGTCCC 1620
QY 1492 TCTTCAAGCTGGCGGAGTGAATTGGCAAGGTGGCAAGACCTGTAAAGCAATTGCAAG 1551
DB 1621 TCTTCAAGCTGGCGGAGTGAATTGGCAAGGTGGCAAGACCTGTAAAGCAATTGCAAG 1680
QY 1552 TTAACAGTGCAGAAATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1611
DB 1681 TTAACAGTGCAGAAATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
QY 1612 ATCTGAGAAATTTATCGGCACTTCTTGTGAGCAGACTGCAAGCGCAGATCAGGGA 1671
DB 1741 ATCTGAGAAATTTATCGGCACTTCTTGTGAGCAGACTGCAAGCGCAGATCAGGGA 1800
QY 1672 ATGTACAAACAGGTGAAGCAGAGAGAGAAATPCCCTCAGGAGTCCCTCAGAGGC 1731
DB 1801 ATGTACAAACAGGTGAAGCAGAGAGAGAAATPCCCTCAGGAGTCCCTCAGAGGC 1860
QY 1732 AGCAAGTGAAGCTCCACAGGACACAGCAAAACAGAGTAAGTAACTGACCTTCCAAAC 1791
DB 1861 AGCAAGTGAAGCTCCACAGGACACAGCAAAACAGAGTAAGTAACTGACCTTCCAAAC 1920
QY 1792 CTGACAGATGAGACCAACCGCAGCCAGATCGGAGCAACCAAGACATCTGAG 1851
DB 1921 CTGACAGATGAGACCAACCGCAGCCAGATCGGAGCAACCAAGACATCTGAG 1980
QY 1852 GAAATGAGAGTCTGCGAGAGGCGGCAAGGACTCTGCCGAGGCTGTGAGAACCCAGGGGC 1911
DB 1981 GAAATGAGAGTCTGCGAGAGGCGGCAAGGACTCTGCCGAGGCTGTGAGAACCCAGGGGC 2040
QY 1912 CGAGAGAGGCGGAGAGTCAAGCAGGTTTGCAGAACCAACGAGCCGCTCCGCGC 1971
DB 2041 CGAGAGAGGCGGAGAGTCAAGCAGGTTTGCAGAACCAACGAGCCGCTCCGCGC 2100
QY 1972 CCCCAGGCTTCTGAGGCTTCAAGCATTCATTCACATTCACATTCGAGTCTCTGAA 2031
DB 2101 CCCCAGGCTTCTGAGGCTTCAAGCATTCATTCACATTCACATTCGAGTCTCTGAA 2160
QY 2032 CTCCACAGAGCTATCCCTTTTGTGTAACCTAATGAGTGAAGCTGTCAAGCCAGC 2091
DB 2161 CTCCACAGAGCTATCCCTTTTGTGTAACCTAATGAGTGAAGCTGTGTCAAGCCAGC 2220
QY 2092 AAAATGACACCCCTTTTCTGTGCAAAATGCTCTGTACATGTGTATTAATAA 2151
DB 2221 AAAATGACACCCCTTTTCTGTGCAAAATGCTCTGTACATGTGTATTAATAA 2280
QY 2152 GGGAGATGTTAAGATATGTGCGCTGTGGTTACAGAGGCTGCGAGCGATATAT 2211
DB 2281 GGGAGATGTTAAGATATGTGCGCTGTGGTTACAGAGGCTGCGAGCGATATAT 2340
QY 2212 TTTAGAAATATATATCAATTAATCAACCTCAACTCCAAATTTTAAATTAATTT 2271
DB 2341 TTTAGAAATATATATCAATTAATCAACCTCAACTCCAAATTTTAAATTAATTT 2400
QY 2272 TTTTCTTTTAAAGAAAGAGGCTTTCTAGACTTTTAAAGAAATTAAGCTTTGGAG 2331
DB 2401 TTTTCTTTTAAAGAAAGAGGCTTTCTAGACTTTTAAAGAAATTAAGCTTTGGAG 2460
QY 2332 GTCTACGCTGTAGAGAGGCTTTTGAAGCCACCCGACAAATTCACCCAGAGGAAT 2391
DB 2461 GTCTACGCTGTAGAGAGGCTTTTGAAGCCACCCGACAAATTCACCCAGAGGAAT 2520
QY 2392 CTGCTGGAAGAGCACTCAGCGGAGTTCTGATCACTGTGTATGTCAACGAAGGATA 2451

Db 2521 CTGCTGGAGAGACACTCAGCGAGTTCTGTGATCACCCTGTGTATGTCAACGAGGATTA 2580
 QY 2452 CCGTCTCTTGAAGAGAACTCTGTCACTCTCTCAATGCTGTCTAAGCTCAATACCCATT 2511
 Db 2581 CCGTCTCTTGAAGAGAACTCTGTCACTCTCTCAATGCTGTCTAAGCTCAATACCCATT 2640
 QY 2512 TCTCTGTGCTTCAAGGTTTAAATGTTTTTTCATAGCTGCTTATATATCTCTGCT 2571
 Db 2641 TCTCTGTGCTTCAAGGTTTAAATGTTTTTTCATAGCTGCTTATATATCTCTGCT 2700
 QY 2572 CTCTGTGTTATCTCTCCCTCCCTCCCTCCCTCTCTCTCTCAATCTCAATCTTTTGA 2631
 Db 2701 CTCTGTGTTATCTCTCCCTCCCTCCCTCCCTCTCTCTCTCAATCTCAATCTTTTGA 2760
 QY 2632 TTTCTCTCAATCTCTCAATCTCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2691
 Db 2761 TTTCTCTCAATCTCTCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2820
 QY 2692 GTGCTCTGATATCATCATCAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2751
 Db 2821 GTGCTCTGATATCATCATCAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880
 QY 2752 CTTCACCTGTTGTTACTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2811
 Db 2881 CTTCACCTGTTGTTACTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2940
 QY 2812 AGGAG 2871
 Db 2941 AGGAG 3000
 QY 2872 AAG 2931
 Db 3001 AAG 3060
 QY 2932 ATTCCTCTCTTAAAAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2991
 Db 3061 ATTCCTCTCTTAAAAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3120
 QY 2992 CAGGAGGTTAAATTCACAGATTTTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAG 3051
 Db 3121 CAGGAGGTTAAATTCACAGATTTTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAG 3180
 QY 3052 GATGTTTAACTGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3111
 Db 3181 GATGTTTAACTGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3240
 QY 3112 TTGGAGCATTTTTTAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3171
 Db 3241 TTGGAGCATTTTTTAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300
 QY 3172 GCCAGCTTGAAG 3231
 Db 3301 GCCAGCTTGAAG 3360
 QY 3232 AGCAG 3283
 Db 3361 AGCAG 3412

RESULT 8
 BC021290 3633 bp mRNA linear PRI 04-OCT-2003
 LOCUS Homo sapiens IGF-II mRNA-binding protein 2, mRNA (cDNA clone
 DEFINITION MGC:29539 IMAGE:5090334), complete cds.
 ACCESSION BC021290
 VERSION BC021290.2 GI:33878041
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens (human); Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 3633)
 AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,

TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE JOURNAL

REMARK COMMENT

FEATURES source

gene
 CDS
 1..3633
 /gene="IGF-II"
 /note="synonym: VICKZ2"
 /db_xref="Locustid:10644"
 68..1864
 /codon_start=1
 /product="IGF-II mRNA-binding protein 2"
 /protein_id="AAH21290.1"

Klausner,R.D., Collins,F.S., Wagner,L., Sherman,C.M., Schuler,G.D.,
 Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhak,N.K.,
 Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
 Ditchenko,L., Marins,K., Farmer,A.A., Rubin,G.M., Hong,L.,
 Stappleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
 Schetz,T.E., Brownstein,M.J., Udell,T.B., Toshitsuki,S.,
 Carninci,P., Prange,C., Raha,S.S., Iqbalilano,N.A., Peters,G.J.,
 Abramson,R.D., Mullany,S.J., Bosak,S.A., McKwan,P.J.,
 McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
 Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Halys,S.W.,
 Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
 Faney,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
 Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
 Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
 Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmitz,J., Myers,R.M.,
 Butlerfield,Y.S., Krzyzanski,M.I., Skalska,U., Smallue,D.E.,
 Schermer,A., Schein,J.E., Jones,S.J., and Marra,M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932
 2 (bases 1 to 3633)
 Strausberg,R.
 Direct Submission
 Submitted (14-JAN-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgs.nci.nih.gov>
 On Aug 19, 2003 this sequence version replaced gi:18204200.
 Contact: MGC help desk
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@nih.gov
 Akheri,N., Aylee,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
 Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
 Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
 Hansen,N., Ho,S.-H., Karlins,E., Kwong,P., Latic,P., Legaapi,R.,
 Maduro,Q.L., Mastello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
 McDowell,J., Pearson,R., Stantropop,S., Thomas,P.J., Touchman,J.W.,
 Tsugeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
 Young,A., Zhang,L.-H. and Green,B.D.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAL Plate: 39 Row: 0 Column: 21
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 27552765.
 Location/Qualifiers
 1..3633
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:29539 IMAGE:5090334"
 /tissue_type="Pancreas, epithelioid carcinoma"
 /clone_lib="NIH_MGC_42"
 /lab_host="DH10B-R"
 /note="Vector: POTB7"
 1..3633
 /gene="IGF-II"
 /note="synonym: VICKZ2"
 /db_xref="Locustid:10644"
 68..1864
 /codon_start=1
 /product="IGF-II mRNA-binding protein 2"
 /protein_id="AAH21290.1"

Db 1741 CAGAAATTAACGGGCACTTTCTGACGACACTGACAGCGCAAGATCAGGAAATTTGT 1800
QY 1677 ACAACAGGTGAAGACAGACAGACAGAAATACCTTCAGGAGTCCGCTTCACAGCCGACAA 1736
Db 1801 ACAACAGGTGAAGACAGACAGACAGAAATACCTTCAGGAGTCCGCTTCACAGCCGACAA 1860
QY 1737 GTGAGGCTCCCAACAGGACACAGCAAAACAGAGGAAATGTAGGCTTCCAAACACTGAC 1796
Db 1861 GTGAGGCTCCCAACAGGACACAGCAAAACAGAGGAAATGTAGGCTTCCAAACACTGAC 1920
QY 1797 AGAATGACACCAACAGCAGCCAGACAGCCAGAGCCAGAAACCAAGACCACTGAGAAATG 1856
Db 1921 AGAATGACACCAACAGCAGCCAGACAGCCAGAGCCAGAAACCAAGACCACTGAGAAATG 1980
QY 1857 AGAAGTCTGCGAGGCGCGCAGAGGACTCTGCGGAGCCCTTGAAACCCCAAGGCGCGAGG 1916
Db 1981 AGAAGTCTGCGAGGCGCGCAGAGGACTCTGCGGAGCCCTTGAAACCCCAAGGCGCGAGG 2040
QY 1917 AGGGGCGGGGAAAGTCAAGCAGAGTTTGCAGAAACAGAGCCCGCTCCGCGCCCA 1976
Db 2041 AGGGGCGGGGAAAGTCAAGCAGAGTTTGCAGAAACAGAGCCCGCTCCGCGCCCA 2100
QY 1977 GGGCTTCTGACAGGCTTCAGCCATCCACTTCAACATCCACTCGAATCTCTGAACTCCG 2036
Db 2101 GGGCTTCTGACAGGCTTCAGCCATCCACTTCAACATCCACTCGAATCTCTGAACTCCG 2160
QY 2037 ACAGCGCTATCCCTTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAA 2096
Db 2161 ACAGCGCTATCCCTTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAA 2220
QY 2097 GGAACCCCTTTTCTGTCGCAATGCTCTGTAACATGCTGTAACATGCTGTAACATGCTG 2156
Db 2221 GGAACCCCTTTTCTGTCGCAATGCTCTGTAACATGCTGTAACATGCTGTAACATGCTG 2280
QY 2157 GATGTTAAGATATGTCGCTGTCGCTGTCGCTGTCGCTGTCGCTGTCGCTGTCGCTGTC 2216
Db 2281 GATGTTAAGATATGTCGCTGTCGCTGTCGCTGTCGCTGTCGCTGTCGCTGTCGCTGTC 2340
QY 2217 AATTAATATATCAATTAATCACTCAATCACTCAATCACTCAATCACTCAATCACTCAAT 2276
Db 2341 AATTAATATATCAATTAATCACTCAATCACTCAATCACTCAATCACTCAATCACTCAAT 2400
QY 2277 CTTTAAAGAGAAACAGGCTTCTGACACTTAAAGATTAAGTCTTTGGAGGCTTC 2336
Db 2401 CTTTAAAGAGAAACAGGCTTCTGACACTTAAAGATTAAGTCTTTGGAGGCTTC 2460
QY 2337 ACAGGTGACAGAGAGCTTTGAGCCACCCGACAAATTCACCCAGAGGAAATCTCGT 2396
Db 2461 ACAGGTGACAGAGAGCTTTGAGCCACCCGACAAATTCACCCAGAGGAAATCTCGT 2520
QY 2397 CGGAAGGACCTACCGGAGTTTGTGATCACTGTGATGTCAGAGAGGATACCGTGC 2456
Db 2521 CGGAAGGACCTACCGGAGTTTGTGATCACTGTGATGTCAGAGAGGATACCGTGC 2580
QY 2457 TCCCTTGAAGAGAACTCTGTCACCTGTCGCTGTCGCTGTCGCTGTCGCTGTCGCTGTC 2516
Db 2581 TCCCTTGAAGAGAACTCTGTCACCTGTCGCTGTCGCTGTCGCTGTCGCTGTCGCTGTC 2640
QY 2517 TTGCTTCAAGGTTTAACTGTTTAACTGTTTAACTGTTTAACTGTTTAACTGTTTAACTGTT 2576
Db 2641 TTGCTTCAAGGTTTAACTGTTTAACTGTTTAACTGTTTAACTGTTTAACTGTTTAACTGTT 2700
QY 2577 TGTATTATCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2636
Db 2701 TGTATTATCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2760
QY 2637 TCAATCCCTCATCTCAATCCCTCATCTCAATCCCTCATCTCAATCCCTCATCTCAATCCCT 2696
Db 2761 TCAATCCCTCATCTCAATCCCTCATCTCAATCCCTCATCTCAATCCCTCATCTCAATCCCT 2819
QY 2697 CTGAGTATCATCATCAACAAAGAGACAAAGGCAAAACACAGAGCTCAACTTAC 2756
Db 2820 CTGAGTATCATCATCAACAAAGAGACAAAGGCAAAACACAGAGCTCAACTTAC 2879

QY 2757 ACTTGTTACTCAAAAGACAGAGTCAATGTACTTGCTCTAGCGTTTGGAGAGAA 2816
Db 2880 ACTTGTTACTCAAAAGACAGAGTCAATGTACTTGCTCTAGCGTTTGGAGAGAA 2939
QY 2817 AACAGAAACCAACCAACCAACCAATCAATCAACCAACCAACCAATCAATCAATCAATCA 2876
Db 2940 AACAGAAACCAACCAACCAACCAATCAATCAACCAACCAACCAATCAATCAATCAATCA 2999
QY 2877 ATGATATTTGCTTTTTCATTTTGTGTATTAAGCATCAATATTCAGCAAAATGATTC 2936
Db 3000 ATGATATTTGCTTTTTCATTTTGTGTATTAAGCATCAATATTCAGCAAAATGATTC 3059
QY 2937 TTTCTTTT-AAAAAAAAAATGTGAGAGAGTAGAAATTTTACAAAGTTTGGCCAG 2995
Db 3060 TTTCTTTTAAAAAAAAAATGTGAGAGAGTAGAAATTTTACAAAGTTTGGCCAG 3119
QY 2996 GCGTTAAATTCACAGATTTTAAACAGAAACCAACAGAGAGAGTCACTCAGGTG 3055
Db 3120 GCGTTAAATTCACAGATTTTAAACAGAAACCAACAGAGAGAGTCACTCAGGTG 3179
QY 3056 TTTTACCTCAGACCTTGCTCTGTTGCTTCCCTTGAAGATTTTGAAGCTGATAGTTG 3115
Db 3180 TTTTACCTCAGACCTTGCTCTGTTGCTTCCCTTGAAGATTTTGAAGCTGATAGTTG 3239
QY 3116 AGCATTTTATTTTATTTTATTTTAAATATGATTTGAAAAAATTAATATCACTGCCA 3175
Db 3240 AGCATTTTATTTTATTTTATTTTAAATATGATTTGAAAAAATTAATATCACTGCCA 3299
QY 3176 GCGTGAAGAGGTGACAGTCCAGAGTGCACAGAGTGTCTGAATGCTCCGCTAGCC 3235
Db 3300 GCGTGAAGAGGTGACAGTCCAGAGTGCACAGAGTGTCTGAATGCTCCGCTAGCC 3359
QY 3236 AAGAACCAATATGAGCTTCTTTTGAACAAACCTTGAATGTTATTT 3283
Db 3360 AAGAACCAATATGAGCTTCTTTTGAACAAACCTTGAATGTTATTT 3407

RESULT 9
AC020629 142971 bp DNA linear PRI 07-MAR-2002
LOCUS Homo sapiens 12q BAC Rpl1-7E16 (Roswell Park Cancer Institute
DEFINITION Human BAC library) complete sequence.
ACCESSION AC020629
VERSION AC020629.6 GI:765675
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 142971)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Bodley,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
Bodley,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferrigno,D.,
Forcum-Tansey,J., Frantz,P., Ganesu,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,M., Harris,K., Hernandez,J., Hodgson,A., Hognes,M.,
Holloway,C., Hosak,H., Issar,A., Jackson,L.E., Jackson,L., Jia,Y.,
Jones,M., Kelly,S., Kneitz,S., Kondeljevski,N., Kong,Y., Kovar,C.,
Lau,S., Leal,B., Lee,E., Li,Z., Lichtarge,O., Liu,J., Liu,M.,
Logan,O., Lozano,R.J., Lu,J., Lucier,R., Marondel,I., Martin,R.,
Martinez,C., McLeod,M.P., Mel,G., Merscher,S., Miller,A.,
Montgomery,K.T., Morgan,M., Morris,S., Naah,S., Nelson,A.,
Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B., Paxton,S.,
Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,D., Rives,M.,
Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H., Shim,C., Simon,M.,
Sparks,A., Stamps,A., Sugang,R., Tabot,P., Taylor,T., Vasquez,L.,
Vanson,R., Vo,O., Wahbah,M., Watlington,S., Weinstein,G.,
Weinstock,I.R., Williamson,A., Worley,K., Wren,J., Wrenford,G.,
Gibbs,A.M., Yang,R., Yu,W., Zhou,X., Kuchelapatti,R., Nelson,D. and
Direct Submission

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 142971)
 AUTHORS Worley,K.C.
 TITLE Direct Submission
 JOURNAL Submitted (07-JAN-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 REFERENCE 3 (bases 1 to 142971)
 AUTHORS Worley,K.C.
 TITLE Direct Submission
 JOURNAL Submitted (27-APR-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 REFERENCE 4 (bases 1 to 142971)
 AUTHORS Worley,K.C.
 TITLE Direct Submission
 JOURNAL Submitted (28-APR-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 REFERENCE 5 (bases 1 to 142971)
 AUTHORS Worley,K.C.
 TITLE Direct Submission
 JOURNAL Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Apr 27, 2000 this sequence version replaced gi:7025656.
 COMMENT INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
 gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
 entire insert of this clone. Overlapping regions of clones are only
 sequenced and submitted once, so the sequence for the remainder of
 the insert may be found in the record for the adjacent clones.
 Overlapping clones are noted at the beginning and end of the
 Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches
 of a local database that includes entries from dbSTS, GDB, and
 local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,
 unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-14) to the
 EST and cDNA sequences. Genes demonstrate at least two exons
 flanked by consensus splice sites that maintained sequence
 continuity across the splice junctions. Sequences that are not
 identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
 standard of double strand coverage with a minimum of 2 clones and 2
 reads with no ambiguities or 2 chemistries with a minimum of 2
 clones and 3 reads with no ambiguities. If the sequence quality for
 a region does not meet this standard, it will be indicated in the
 annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
 standards - estimated error rate less than 1 per 10,000 bases.
 Reports of lowest quality individual bases and measures of base
 quality are listed below. Description of the metrics can be found
 at URL:
<http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics -----
 Contig length: 142971
 Phrap values in estimate: 140680
 Average error rate (BCM-Phrap estimate): 0.000118703
 Fraction of Phrap values less than 40 : 0.0433608
 Number of consensus changing edits: 31
 Number of N's in consensus : 0

Position	Original+Context	Edited+Context
11171	gtctctcga (n) caaaaaaa	gtctctcga (a) caaaaaaa
12018	tatctatct (n) ctatctatc	tatctatct (c) ctatctatc
16604	tcacacctaa (n) agttccatt	tcacacctaa (g) agttccatt
18385	tggtaaccc (n) taccatttca	tggtaaccc (t) taccatttca
21879	aaaaaaaaa (n) ntttaaacaa	aaaaaaaaa (a) ntttaaacaa
21880	aaaaaaaaa (n) atttaaacag	aaaaaaaaa (a) atttaaacag
21881	aaaaaaaaa (n) ttaaacagag	aaaaaaaaa (a) ttaaacagag
21906	ccaggtccc (n) cttaatagag	ccaggtccc (a) cttaatagag
38215	accagctcgt (n) caaatgycg	accagctcgt (c) caaatgycg
55675	ttacatctag (n) tcttctgcaa	ttacatctag (t) tcttctgcaa
55676	acttctctc (n) aaaaaaa	acttctctc (a) aaaaaaa
74371	gaccagctg (n) ccaacatcgt	gaccagctg (a) ccaacatcgt
88021	gtctcgccc (n) ccgcccgcg	gtctcgccc (c) ccgcccgcg
88248	taaatgagg (n) agaccatctg	taaatgagg (c) agaccatctg
94121	ggaatgagcc (n) agatcacacc	ggaatgagcc (a) agatcacacc
95228	ttgtccctga (n) gcttcctcgc	ttgtccctga (t) gcttcctcgc
95230	gtccatggtc (n) cmnttctgc	gtccatggtc (t) cmnttctgc
95231	ccatgtgttc (n) nttgttcaa	ccatgtgttc (t) nttgttcaa
95232	catgtgttc (n) ntgttcaac	catgtgttc (c) ntgttcaac
104931	atgtgttctc (n) ttgttcaact	atgtgttctc (a) ttgttcaact
118168	tttttttag (n) acgaacagg	tttttttag (t) acgaacagg
120974	gcccaagtc (n) ccaaatggtc	gcccaagtc (c) ccaaatggtc
120976	catctcaaa (n) ananagaatc	catctcaaa (a) ananagaatc
120978	tcctcaaaa (n) anagaatctc	tcctcaaaa (a) anagaatctc
129505	caaanana (n) gaactctcc	caaanana (a) gaactctcc
135050	ccagccatgg (n) gnyggcacc	ccagccatgg (t) gnyggcacc
135058	ggcatgggg (n) gggcaccgtc	ggcatgggg (t) gggcaccgtc
139698	attacatca (n) aatatgtag	attacatca (g) aatatgtag
140104	accgtgccca (n) cagaatgaa	accgtgccca (g) cagaatgaa
140149	tggaaccatt (n) gataccocaa	tggaaccatt (g) gataccocaa

----- Distribution of Quality < 40 Bases -----

#	5	10	15	20	25	30	35	40
1000	*	*	*	*	*	*	*	*
900	*	*	*	*	*	*	*	*
800	*	*	*	*	*	*	*	*
700	*	*	*	*	*	*	*	*
600	*	*	*	*	*	*	*	*
500	*	*	*	*	*	*	*	*
400	*	*	*	*	*	*	*	*
300	*	*	*	*	*	*	*	*
200	*	*	*	*	*	*	*	*
100	*	*	*	*	*	*	*	*
0	*	*	*	*	*	*	*	*

FEATURES

Version: 1.01 gxt0.
 Location/Qualifiers
 source 1..142971
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="12"
 /clone="RP11-76E16"
 repeat_region 147..236
 /rpt_family="A-rich"
 repeat_region 258..378
 /rpt_family="AluB"
 repeat_region complement(537..840)
 /rpt_family="AluX"
 repeat_region complement(1722..2137)
 /rpt_family="Tiger2a"
 repeat_region complement(3161..3584)

		/rpt_family="MSTD"	
STS		3228..3502	
		/standard_name="G05956"	
STS		/db_xref="dbSTS:11453"	
		3347..3490	
		/standard_name="CdaOybo4"	
		/db_xref="GDB:446196"	
		3585..3735	
misc_feature		/note="Region similar to Hs#S936143 aj27402.s1 Homo sapiens cDNA: AA789222"	
repeat_region		3885..3910	
		/rpt_family="(TA)n"	
repeat_region		complement(3917..4030)	
repeat_region		/rpt_family="Alubd"	
		complement(4544..4681)	
misc_feature		5145..5510	
		/note="Region similar to Hs#S93417 ca06g05.s1 Homo sapiens cDNA: AA765439"	
misc_feature		5533..5859	
		/note="Region similar to Hs#S1195593 qt69f10.x1 Homo	
Query Match	85.3%	Score 2799.4;	DB 9; Length 14297;
Best Local Similarity	93.7%	Pred. No. 0;	Mismatches 157; Indels 51; Gaps 14;
Matches 3084;	Conservative	0;	
OY	1	GGCAGCGGAGGAGGAGAGACGCCGCGGTATCCGGCCGGGAGGCGCGGCTCTCG393G	60
Db	132954	GGCAGCAGAGAGAGGAGAGGCGGAGCGCATTCGGG-CGGGGAGACCGCGGCTCG393G	133012
OY	61	AAGAAGCGAATGATATACAAGCTTTACATCGGGAACCTGAGCCCGCGTCAACGCGCAG	120
Db	133013	AACAGACGGAATGATATAACAGCTTTTACTCGGAACTGAGCCCGCGTCAACGCGCGAA	133072
OY	121	GACCCTCGGACAGCTCTTGCGGAGACAGAACTGCCCTCGCGCGGACAGGTCTCTGCTGAAG	180
Db	133073	GACCTCGGACAGCTCTTGCGGAGACAGAACTGCCCTCGCGGACAGGTCTCTGCTGAAG	133132
OY	181	TCCGGCTACGCTTCTGTGACTATCCCCGACACAAACTGGGCGCATCCGCGCCATTCGAGACC	240
Db	133133	TCCCCTACGCTTCTGTGACTATCCCCGACACAAACTGGGCGCATCCGACATCGAGACC	133192
OY	241	CTCTCGGTAAGATGGAATTGCATGGGAAAAATCATGAAAGTTGATTCTCACTCTCTAAA	300
Db	133193	CTCTCGGCTCAATGGAATTGCATGGGAAAAATCATGAAAGTTGATTCTCACTCTCTAAA	133252
OY	301	AAGCTTAGAGACGAGAAAATTCAGATTGGAATCCTCTCTCACTCGAGTGGAGGTG	360
Db	133253	AAGCTTAGAGACGAGAACCTTCGATTGGAATATCCCTCTCACTCGAGTGGAGGTG	133312
OY	361	TTGATATGACTTTTGGCTCAATATGGGACAGTGGGAAATGTGGAAACAATCAAACAGAC	420
Db	133313	TTGATATGACTTTTGGCTCAATATGGGACAGTGGGAAATGTGGAAACAATCAAACAGAC	133372
OY	421	ACAGAAACCGCGTGTCAACGTCAATATGCAACAAGAGAAACAAAATATGCCATG	480
Db	133373	ACAGAACCGCGTGTGTCAACGTCAATATGCAACAAGAGAAACAAAATATGCCATG	133432
OY	481	GAGAACTTAAGCGGCGCATCAGTTTGAAGTACTCTCTTCAAGATTTCTTCAATCCCGAT	540
Db	133433	AAGAACTTAAGCGGCGCATCAGTTTGAAGTACTCTTCAAGATTTCTTCAATCCCGAT	133492
OY	541	GAAAGGATAGCTCCCTTCGCCCTCACGACGACGAGTGGGAGACACTTTCOCGG	600
Db	133493	GACGAGTACGCTGCCCTTCGCCCTCACGACGACGAGTGGGAGACACTTTCOTGG	133552
OY	601	GAGCAAGGACGCGCCCTGGGGGACTTTCAGGCGACAGATTGATTTCCCGCTGCGG	660
Db	133553	GAGCAAGGACGAGCCCTGGGGGACTTTCAGGCGACAGATTGATTTCCCACTGCGT	133612
OY	661	ATCTCGTCCCAACCAAGTTTGTGTGTCATCATGGAAGGAGGGCTTGAACATTAAG	720
Db	133613	ATCTCGTCCCAACCAAGTTTGTGTGTCATCATGGAAGGAGGGCTTGAACATTAAG	133672

QY	721	AACATCACTAAGCAGACCCAGTCCCGGATGATATTCATAGAAAAGAACTCTGAGCT	780
Db	133673	AACATCACTAAGCAGAGCCGGTCCCGGATGATATTCATAGAAAAGAACTCTGAGCT	133732
QY	781	GCAGAGAAAGCTGTGCACCTCATGTCACCCCGAGGGGACTTCTGAAAGCATGCGGATG	840
Db	133733	GCAGAGAAAGCTGTGCA-----CGATGCCACCCAGAGGGGACTTCTGAAAGCATGCGGATG	133788
QY	841	ATTCTGGAATCATGCGAAAGAGAGCATGAGACCAACTAGCCGAGAGATTCTCTG	900
Db	133789	ATTCTGGAATTAATGCGAAAGAGAGCATGAGCCAACTAGCGAGAGATTCTCTG	133848
QY	901	AAAACTTTGGCACAACAATGCTGTGGTGAAGTGTATGGAAAAAGAAAGCAAAATTGG	960
Db	133849	AAAACTTTGGCCCAATGAGCTGTGGTGAAGTGTATGGAAAAAGAAAGCAAAATTGG	133908
QY	961	AAGAAAAATTGAACATGAAAACAGGGAACCAAGTAACATCTCATCTTGGCAGATTGAGC	1020
Db	133909	AAGAAAAATTGAACATGAAAACAGGGAACCAAGTAACATCTCATCTTGGCAGATTGAGC	133968
QY	1021	ATATACAAACCCGGAAAAGAACCATCATCTGTGAAGGGCAGATTGAGGCTGTGCGAGTCT	1080
Db	133969	ATATACAAACCCGGAAAAGAACCATCATCTGTGAAGGGCAGATGAGGCTGTGCGAGTCT	134028
QY	1081	GAGATAGAGATTAATGAAGAGCTGCGTAGGAGCTTTGAAAATGATATGCTGTGTAAC	1140
Db	134029	GAGATAGAGATTAATGAAGAGCTGCGTAGGAGCTTTGAAAATGATATGCTGTGTAAC	134088
QY	1141	ACCACCTCCGGAATCTTCTCCAGCTGTGACCCCATACACGTTTGGCCGGTCCCGCAT	1200
Db	134089	ACCACCTTCGGATACTTCTCCAGCTGTGACCCCATACGCTTTGGCCGGTCCCGCAT	134148
QY	1201	CATCACTCTTATCCAGAGCAGAGATTGTGAATCTCTTCATCCCAACCCAGAGCTGTGGGC	1260
Db	134149	CATCACTCTTATCCAGAGCAGAGATTGTGAATCTCTTCATCCCAACCCAGAGCTGTGGGC	134208
QY	1261	GCCATCATTCGGGAAGAGGGGACACATCAACAGCTGGCGAGATTGCCGAGCTCT	1320
Db	134209	GCCATCATTCGGGAAGAGGGGACACATCAACAGCTGGCGAGATTGCCGAGCTCTC	134268
QY	1321	ATCAAGATTGCCCTTCGCGAAGGCCCAAGCTCAGCGAAAGGATGTCATCAACCCGG	1380
Db	134269	ATCAAGATTCGCCCTTCGCGTAG-----CGAAGGTCAATCAACCTGG	134311
QY	1381	GCACCGGAAAGCCAGTTCAGAGGCCCGAGGAGCGATCTTTGGGAAACTGAAAGGAAAC	1440
Db	134312	GCACCGGAATCCCACTGTTCAAGGCCAGGAGCGATCTTTGGGAAACTGAAAGGAAAC	134371
QY	1441	TTCTTTAACCCCAAGAAAGAAAGTGAAGCTGGAAGCGCATATCAGAGTCCCTCTTCACA	1500
Db	134372	TTCTTTAACCCCAAGAAAGAAAGTGAAGCTGGAAGCCCATATCAGAGTCCCTCTTCACC	134431
QY	1501	GCTGGCCCGGATGATTGGCAAAAGTGGCAAGACCGTAAGCACTGAGAACTTAACAGT	1560
Db	134432	GCTGGCCCGGATATTGGCAAAAGGAGCAAGCTGTAATGAACCTGAGAACTTAATCACT	134491
QY	1561	GCAGAAAGTATGCTGCTCGTGAACCAACGCCAGATGAATAATGAGGAAATGATCGTCA	1620
Db	134492	GCAGAAAGTATGCTGCTCGTGAACCAACGCCAGATGAATAATGAGGAAATGATCGTCA	134551
QY	1621	ATTATCGGACACTTCTTTCTAGCCAGACTGCAAGCGCAAGATCAAGGAAATTGTACAA	1680
Db	134552	ATTATCGGACACTTCTTTCTAGCCAGACTGCAAGCGCAAGATCAAGGAAATTGTACAA	134611
QY	1681	CAGGTGAAGCAGAGAGCAAGAAATCCCTCAGGAGTGGCTCAAGCGCAAGCAAGTGA	1740
Db	134612	CAGGTGAAGCAGAGAGCAAGAAATCCCTCAGGAGTGGCTCAAGCGCAAGCAAGTGA	134671
QY	1741	GGCTCCCAAGGACCAAGCAAAACAAGATGAATATGAGCTTTCACAACCTGACAGAA	1800
Db	134672	GGATCCCAAGGACCAAGCAAAACAAGAGATATGAGCTTTCACAACCTGACAGAA	134731

QY 1801 TGAAGCAAAAGCAGCAGCAGATCGGAGCAAAACAAAGACATCTGAGGATGAGA 1860
 DB 134732 TGAAGCAAAAGCAGCAGCAGATCGGAGCAAAACAAAGACATCTGAGGATGAGA 13491
 QY 1861 GTCCTCGAGGCGGCGGAGGATCTTGCAGAGCCCTGAGAACCCGAGGCGGAGAGG 1920
 DB 134792 GTCCTCGAGGCGGCGGAGGATCTTGCAGAGCCCTGAGAACCCGAGGCGGAGAGG 134851
 QY 1921 GCGGGAAGGTGAGCAGGATTTGCGAGAACCAAGAGCCCGCCGCCGCCAGAGG 1980
 DB 134852 GCGGGAAGGTGAGCAGGATTTGCGAGAACCAAGAGCCCGCCGCCGCCAGAGG 134911
 QY 1981 TTTCGAGGCTTCAGCCATCCACTTCACCATCTCGGATCTCTGAACTCCCAAGA 2040
 DB 134912 TTTCGAGGCTTCAGCCATCCACTTCACCATCTCGGATCTCTGAACTCCCAAGA 134971
 QY 2041 CGCATCCCTTTTAACTGAACTACATAGGTGAGC-GTGTCAAAAGCCAGCAAAATGCA 2099
 DB 134972 CGCATCCCTTTTAACTGAACTGAAATAGTAAACATTTGTTCAAGCCAGCAAAATGCA 135031
 QY 2100 CACCCCTTTTCTGAGCAAAATGCTCTGATCATGTGTGATCATTTTGAAGAGGAGAT 2159
 DB 135032 CGCCCTTTTCTGAGCAAAATGCTCTGATCATGTGTGATCATTTTGAAGAGGAGAT 135090
 QY 2160 GTTAAGATATGTGCGCTGTGAGTACACAGGCTGCGAGCGGTAATATTTTGAAGA 2219
 DB 135091 GTTAAGATATGTGCGCTGTGAGTACACAGGCTGCGAGCGGTAATATTTTGAAGA 135150
 QY 2220 TAAATATCAAAATCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 2279
 DB 135151 TAAATATCAAAATCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 135202
 QY 2280 TTTAAGAGAGAGGCTTTCTAGACTTTAAGAAATAAGCTTTTGGAGGCTCAGC 2339
 DB 135203 TTTAAGAGAGAGGCTTTCTAGACTTTAAGAAATAAGCTTTTGGAGGCTCAGC 135259
 QY 2340 GTTAGAGAGAGCTTTGAGGCCACCGCACAATAATTCACAGAGGAAATCTCGCGG 2399
 DB 135260 GTTAGAGAGAGCTTTGAGGCCACCGCACAATAATTCACAGAGGAAATTTTGCAG 135319
 QY 2400 AAGGACACTCAGGCACTGTTGATCAGCTGTGATGTCAACAGAGGATACCGTCTCC 2459
 DB 135320 AAGGACACTCAGGCACTGTTGATCAGCTGTGATGTCAACAGAGGATACCGTCTCC 135379
 QY 2460 TTGAAGAGAGAACTCTGCACTGCTCATGCTGTGATGTCAACAGGATACCGTCTCC 2519
 DB 135380 TTGAAGAGAGAACTCTGCACTGCTCATGCTGTGATGTCAACAGGATACCGTCTCC 135439
 QY 2520 CTTACAGGTTTAACTGTTTTCATAGTCTATTAATTTCTGCTCTCTGT 2579
 DB 135440 CTTACAGGTTTAACTGTTTTCATAGTCTATTAATTTCTGCTCTCTGT 135496
 QY 2580 TTATCTCTCCCTCCCTCCCTCCCTCCCTCTTCTTCATCTCCATCTTTTGAATTTCTCA 2639
 DB 135497 TTATCTCTCCCTCCCTCCCTCCCTCCCTCTTCTTCATCTCCATCTTTTGAATTTCTCA 135555
 QY 2640 TCCCTCATCTCAATCCGATCTGAGGAC-----CCCCCCCCCGAGGAGAAAGCA 2691
 DB 135556 TCCCTCATCTCAATCCGATCTGAGGAC-----CCCCCCCCCGAGGAGAAAGCA 135615
 QY 2692 GTGCTCTGAGTATCATCAACAAAGAGAAAGAAAGCAACAAAGAGGCTCA 2751
 DB 135616 GTGCTCTGAGTATCATCAACAAAGAGAAAGAAAGCAACAAAGAGGCTCA 135675
 QY 2752 CTTACACTTGTGTTACTCAAAAGAGAGTCAATGTTCTTCTGAGGTTTGGAG 2811
 DB 135676 CTTACACTTGTGTTACTCAAAAGAGAGTCAATGTTCTTCTGAGGTTTGGAG 135735
 QY 2812 AGGAAAG 2871
 DB 135736 AGGAAAG 135795
 QY 2872 AAAAGATGATTTTGTCTTTTGTGATTTGTTGATTAAGCATCAATATTCAGCAAAATG 2931

DB 135796 AAAGATGATTTTGTCTTTTGTGATTTGTTGATTAAGCATCAATATTCAGCAAAATG 135855
 QY 2932 ATTCCTTTCTTTAAAAAATGTGAGGAAAGTAGAAATTTA-CCAAGGTTTGTGAC 2990
 DB 135856 ATTCCTTTCTTT-AAAAAATGTGAGGAAAGTAGAAATTTA-CCAAGGTTTGTGAC 135914
 QY 2991 CCAAGGCTTTAAATTCACAGATTTTTTTAAACGAGAAAAACACAGCAG-AGAAGCTACCTC 3049
 DB 135915 CCAAGGCTTTAAATTCACAGATTTTTTTAAACGAGAAAAACACAGCAAGAAAGCTACCTC 135974
 QY 3050 AGTG-TTTTACCTCAGCAGCTGCTGTGTGTGTTCCCTTAAAGATTTTGAAGCTCA 3108
 DB 135975 AGTGTTTAACTCAGCAGCTGCTGTGTGTGTTCCCTTAAAGATTTTGAAGCTCA 136034
 QY 3109 TAGTGGAGCATTTTATTTTAAATTAATAAATGAGTGGAAAAAATAAGATATCA 3168
 DB 136035 CAGTTGGAGCATTTTAAATTTTAAATTAATAAATGAGTGGAAAAAATAAGATATCA 136094
 QY 3169 ACTGCCAGCTGAGAGAGTACAGTCCAGTGTGCAACAGCTGTCTGAATGTCTCC 3228
 DB 136095 ACTGCCAGCTGAGAGAGTACAGTCCAGTGTGCAACAGCTGTCTGAATGTCTCT 136154
 QY 3229 GCTAGCCAGAGAACCNATATGSCCTTCTTTGAGCAAACTTGAAATGTCTTA 3280
 DB 136155 GCTAGCCAGAGAACCNATATGACCTTTTGGCAAACTTTGAAATGTCTTA 136206
 RESULT 10
 AL596177
 LOCUS
 DEFINITION Human DNA sequence from clone Rpl1-325P15 on chromosome 1, complete sequence.
 ACCESSION AL596177
 VERSION AL596177.4 GI:1552942
 KEYWORDS HMG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE
 AUTHORS Clark, S.
 TITLE Direct Submission
 JOURNAL Submitted (08-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 COMMENT
 On Sep 11, 2001 this sequence version replaced gi:15022347.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SwissProt; Tr, TrEMBL; Wp, WormBEP; information on the WormBEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormep
 This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>
 Rpl1-325P15 is from the library RPl1-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone Rp11-325P15. It may be shorter because we sequence overlapping sections only once, except for a short overlap. The true left end of clone Rp11-325P15 is at 1 in this sequence. The true left end of clone Rp11-337C18 is at 98085 in this sequence. The true right end of clone Rp4-704D21 is at 41034 in this sequence.

FEATURES
Location/Qualifiers

1..91084
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/clone="Rp11-325P15"
/clone_11b="RPCT-11.2"

ORIGIN

Query Match 84.7%; Score 2781.4; DB 9; Length 91084;
Best Local Similarity 93.5%; Pred. No. 0;
Matches 3089; Conservative 0; Mismatches 167; Indels 48; Gaps 16;

```

QY 1 GGCACCGAGAGAGGCGCGGCTACCGGCGCGGGAGCCGCGGCTTCGCGG 60
DB 9381 GGCACCGAGAGAGGCGCGGCTACCGGCGCGGGAGCGCGGCTTCGCGG 9440
QY 61 AAGAGACGGATGATGAACAAGCTTTACATCGGGAACTGAGCCCGCGGTCAACGCGGAC 120
DB 9441 AAGAGACGGATGATGAACAAGCTTTACATCGGGAACTGAGCCCGCGGTCAACGCGGAC 9500
QY 121 GACCTCCGCGACGCTCTTGGGAGCAGAAAGCTGCCCTGGCGGAGAGGCTCTGTGAAG 180
DB 9501 GACCTCTGGCACTCTTGGGAGCAGAAAGCTGCCCTGGCGGAGAGGCTCTGTGAAG 9560
QY 181 TCCGCTACGCTTGTGCTGCTACCCCGACAGAACTGGGCGCATCGGCGCATCGAGACC 240
DB 9561 TCCGCTACGCTTGTGCTGCTACCCCGACAGAACTGGGCGCATCGGCGCATCGAGACC 9620
QY 241 CTCTCGGGTAAGTGAATTTGATGGGAAATCATGGAAGTTGATTACTCAAGCTTAA 300
DB 9621 CTCTCGGGTAAGTGAATTTGATGGGAAATCATGGAAGTTGATTACTCAAGCTTAA 9680
QY 301 AAGCTAAGAGAGCAGAAATTCAGATTGCAAAATCCTCTCACTGAGTGGAGGTG 360
DB 9681 AAGCTAAGAGAGCAGAAATTCAGATTGCAAAATCCTCTCACTGAGTGGAGGTG 9740
QY 361 TTGATGACTTTGGCTCAATATGGAAGAGTGGAAATGTGG--AACTGTCACAA 417
DB 9741 TTGATGACTTTGGCTCAATATGGAAGAGTGGAAATGTGGAAATGTGGAAATGTGG 9800
QY 418 GACACGAAACCGCGCTTGTCAACGTCACTATGCAACAAGAGAAAGAAATAGCC 477
DB 9801 GACACGAAACCGCGCTTGTCAACGTCACTATGCAACAAGAGAAAG--AAATAGAC 9855
QY 478 ATGAGAAGCTAAGCGGCGCATCAATTGAGAACTACTCTTCAAGATTTCTTCAATCCG 537
DB 9856 ATGAGAAGCTAAGCGGCGCATCAATTGAGAACTACTCTTCAAGATTTCTTCAATCCG 9915
QY 538 GATGAAGAGTGAAGCTTCCCTTGGCCCTCAAGGAGCCCAAGGTGGGAGCACTCTTC 597
DB 9916 GATGAAGAGTGAAGCTTCCCTTGGCCCTCAAGGAGCCCAAGGTGGGAGCACTCTTC 9975
QY 598 C-GGGAGCAAGGCAAGCGCCCTGGGGGCACTTCTCAGGCGCAGACAGATTATTTCCGCT 656
DB 9976 CGGGAGCAAGGCAAGCGCCCTGGGGGCACTTCTCAGGCGCAGACAGATTATTTCCGCT 10035
QY 657 GCGGATCTGTGCTCCCAAC-----CAGTTGTGTGTCATCATCGAAAGAGGCTT 710
DB 10036 GCGGATCTGTGCTCCCAACAGTTTGTGTGTCATCATCGAAAGAGGCTT 10095
QY 711 GACCATTAAGAAATCATCTAAGCAGACCGAGTCCCGGTAGATTTCCATGAAAAGGAA 770
DB 10096 GACCATTAAGAAATCATCTAAGTGAAGCCAGTCCCGGTAGATTTCCATGAAAAGGAA 10155

```

```

QY 771 CTCTGAGCTGCAGAGAAAGCTGTACCATTCATGCACCCCAAGAGGAGACTTCTGAAGC 830
DB 10156 CTCTGAGCTGCAGAGAAAGCTGTACCATTCATGCACCCCAAGAGGAGACTTCTGAAGC 10215
QY 831 ATGCCGATGATTTCTTGAATTCATGCAAGAGGCGACATGAGCAACACTAGCCGAGAA 890
DB 10216 ATGCCGATGATTTCTTGAATTCATGCAAGAGGCGACATGAGCAACACTAGCCGAGAA 10275
QY 891 GATTCCTCTGAAAATCTTGGGACACAAATGCTTGGTGGAAAGTGAATTTGAAAAGAG 950
DB 10276 GATTCCTCTGAAAATCTTGGGACACAAATGCTTGGTGGAAAGTGAATTTGAAAAGAG 10323
QY 951 CAGAAATTTGAAGAAATTTGAACATGAAACAGGACCAAGATTAACATCTCATCTTGA 1010
DB 10324 CAGAAATTTGAAGAAATTTGAACATGAAACAGGACCAAGATTAACATCTCATCTTGA 10383
QY 1011 GATTTTGAACATTAACAACCGGAAAGAACATCACTGTGAAGGCGACAG--TTGAGGCT 1069
DB 10384 GATTTTGAACATTAACAACCGGAAAGAACATCACTGTGAAGGCGACAGTTTGAAGGCT 10443
QY 1070 GTGCCAGTGTGAGATGAGATTAAGAGAAAGCTGCTGAGGCTTTGAAAATGATATGC 1129
DB 10444 GTGCCAGTGTGAGATGAGATTAAGAGAAAGCTGCTGAGGCTTTGAAAATGATATGC 10503
QY 1130 TGGCTGTAAACCCATCTCCGATCTCTTCCAGCTGTACCCCATCAACGATTGAGC 1189
DB 10504 TGGCTGTAAATACGACTCCGATCTCTTCCAGCTGTACCCCATCAACGATTGAGC 10563
QY 1190 CGTTCCCGATATACATCTTATCCAGAGCAGAGATTGTGATCTCTTATCCCAACC 1249
DB 10564 CGTTCCCGATATACATCTTATCCAGAGCAGAGATTGTGATCTCTTATCCCAACC 10623
QY 1250 AAGCTGTGGCGCATCATCGGAAAGAGGCGCACATCAAAACAGCTGGGAGATTGG 1309
DB 10624 AAGCTGTGGCGCATCATCGGAAAGAGGCGCACATCAAAACAGCTGGGAGATTGG 10683
QY 1310 CCGAGACTCTATCAAGATTTGCTGCGGAGAGCCAGACGTCAGGAAAGATGTGA 1369
DB 10684 CCGAGACTCTATCAAGATTTGCTGCGGAGAGCCAGACGTCAGGAAAGATGTGA 10743
QY 1370 TCATCAACCGGAGCCACCGAGAGCCAGTTCAAGGCCAGGAGCGGATCTTGGGAACTGA 1429
DB 10744 TCATCAACCGGAGCCACCGAGAGCCAGTTCAAGGCCAGGAGCGGATCTTGGGAACTGA 10803
QY 1430 AAGAGAAAATCTTTTAAACCCCAAGAAAGTGAAGCTGGAAGCATATCAAGTGC 1489
DB 10804 AAGAGAAAATCTTTTAAACCCCAAGAAAGTGAAGCTGGAAGCATATCAAGTGC 10863
QY 1490 CCTCTTCCACAGCTGGCGCGGTGATTTGGCAAGAGTGGCAACCGTGAACGACTGAGA 1549
DB 10864 CCTCTTCCACAGCTGGCGCGGTGATTTGGCAAGAGTGGCAACCGTGAACGACTGAGA 10923
QY 1550 ACTTAACAGTGCAGAAATCATCTGCTCTGTGACCAACGCCAGATGAATAAGAGAG 1609
DB 10924 ACTTAACAGTGCAGAAATCATCTGCTCTGTGACCAACGCCAGATGAATAAGAGAG 10983
QY 1610 TGAATGCAAGATTAATCGGCACTTCTTGTGACCAAGCTGCAACAGCGGAATCAAGG 1669
DB 10984 TGAATGCAAGATTAATCGGCACTTCTTGTGACCAAGCTGCAACAGCGGAATCAAGG 11043
QY 1670 AAATTTGACAAAGGTGAAGCAGACAGAGCAAGAAATTAACCTGAGGAGTGCCTCAGC 1729
DB 11044 AAATTTGACAAAGGTGAAGCAGACAGAGCAAGAAATTAACCTGAGGAGTGCCTCAGC 11103
QY 1730 GCAGCAAGTGAAGCTCCACAGGACACAGCAAAACAGATGAATGAGACCTTCCAC 1789
DB 11104 GCAGCAAGTGAAGCTCCACAGGACACAGCAAAACAGATGAATGAGACCTTCCAC 11163
QY 1790 ACCTGACAGAAATGAGCAAAACGACGACGATGAGGAGCAACCAAGACCATCTG 1849
DB 11164 ACCTGACAGAAATGAGCAAAACGACGACGATGAGGAGCAACCAAGACCATCTG 11223
QY 1850 AGGATGAGAAATCTGGGAGGCGGCGAGGAGACTCTGCGGAGGCGCTGAGAAACCCAGGG 1909

```

Db	12297	CAGCAAAAATGATTTCTTTCTTT--AAAAAAAATGTGGAGAAAGTAGAAATTTTACCAGG	12355
Qy	2982	GTGTGTTGGCCACGAGGCGTTAAATTCACAGATTTTTTTTAAACGAGAAAAACACACAGAGAA	3041
Db	12354	GTGTCTGGCCACGAGGCAATTAAATTACAGATTTTTTTTAAATGGAGAAAAACACACAGAAAAA	12413
Qy	3042	GCTAAGCTAGAGTG- TTTTAACTGACGACCTGCTCTGTGTTCCTTAGAGATTTTGT	3100
Db	12414	GCTGAGCTCAGAGTGTTTTTTACCTCAGACCTGCTCTGTGTTCCTTAGAGATTTTGT	12472
Qy	3101	AAAGCTGATGTGTGGAGCATTTTTTTTA-TTTTTTAAATAAAATGAGTTGGAAAAAAAT	3159
Db	12474	AAAACTGATGTGTGGAGCATTTTTTTTAATTTTTTAAATAAAATGAGTTGGAAAAAAAT	12533
Qy	3160	AAGATATCAACTGCGACGACCTGGAGAAAGTGACAGCTCAGATGTGGAAACAGCTGTTGAA	3219
Db	12534	AAGATATCAACTGCGACGACCTGGAGAAAGTGACAGCTCAGATGTGGAAACAGCTGTTGAA	12593
Qy	3220	TGTCTTCGCTGACGACGAGAACCMATATGAGCTTCTTTTGGACAAACCTTGAATGTT	3279
Db	12594	TGTCTTCGCTGACGACGAGAACCMATATGAGCTTCTTTTGGACAAACCTTGAATGTT	12655
Qy	3280	ATTT 3283	
Db	12654	ATTT 12657	
RESULT 11			
AC015706			
LOCUS	AC015706	182695 bp	DNA linear HTG 24-AUG-2002
DEFINITION	Homo sapiens chromosome 1 clone RP11-325P15 map 1, WORKING DRAFT		
ACCESSION	AC015706		
VERSION	AC015706.3		
KEYWORDS	HTG; HTGS PHASE1; HTGS_DRAFT.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiiridae; Homo.		
AUTHORS	Birren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	1 (bases 1 to 182695)		
JOURNAL	Homo sapiens chromosome 1, clone RP11-325P15		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 182695)		
	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barna, N., Becker, R., Boguslavsky, L., Bouckgalter, B., Brown, A., Castle, A., Colangelo, M., Collins, S., Colliore, A., Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kam, L., Karatas, A., Klein, J., Lehoczy, J., Lien, C., Locke, K., Macdonald, P., Marguis, N., McEwan, P., McQuirk, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testage, S., Tirrell, A., Vassiliou, H., Vo, A., Wheeler, J., Wu, X., Wymann, D., Ye, M. J., Zimmer, A. and Zody, M.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		
REFERENCE	3 (bases 1 to 182695)		
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Bouckgalter, B., Brown, A., Burkett, G., Campagnaro, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Colliore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodde, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kam, L., Karatas, A., Klein, J., Labocque, K., Lamazates, R., Lander, E., Lehoczy, J., Levine, R., Lien, C., Liu, G., Liu, G., Locke, K., Macdonald, P., Marguis, N.,		

QY 241 CTCTGGGTAAAGTGAATTCGATGGAAAAATCATGGAAGTTGATTACTCAGTCTCTAA 300
|
|
|
Db 19486 CTCTGGGTAAAGTGAATTCGATGGAAAAATCATGGAAGTTGATTACTCAGTCTCTAA 19545
|
|
|
QY 301 AAGCTAAGGAGGAGAAAAATTCAGATTCGAAACATCCCTCTCAGCTGAGTGGAGGTG 360
|
|
|
Db 19546 AAGCTAAGGAGGAGAAAAATTCAGATTCGAAACATCCCTCTCAGCTGAGTGGAGGTG 19605
|
|
|
QY 361 TTGATGAGCTTTTGGCTCAATATGGAACAGTGGAAATGTGG---AACAGTCAACA 417
|
|
|
Db 19606 TTGATGAGCTTTTGGCTCAATATGGAACAGTGGAAATGTGGAAACAACTCAACA 19665
|
|
|
QY 418 GACACAGAAACCGCGTGTCAACGTCACATATGCAACAGAGAAAGCAAAATATACC 477
|
|
|
Db 19666 GACACAGAAACCGCTGTGTCAACGTCACATATGCAACAGAGAAAG---AAATATAGC 19720
|
|
|
QY 478 ATGAGAAAGCTPAGCGGGCATCAGTTGAGAACTACTCTTCAGATTTCTTACATCCG 537
|
|
|
Db 19721 ATGAGAGAGCTPAGCGGGCATCAGTTGAGAACTAGTCTTCAGATTTCTTACATCCG 19780
|
|
|
QY 538 GATGAAGAGTGAAGTCCCTTCGCCCCCTCAGCGAGCCGACGCTGGGAGCACTCTCC 597
|
|
|
Db 19781 GATGAAGAGTGAAGTCCCTTCGCCCCCTCAGCGAGCCGACGCTGGGAGCACTCTCTC 19840
|
|
|
QY 598 C-GGAGAGCAAGGCGACGCGCCCTGGGGGCACTTCTCAGGSCAGACAGATTTTCCGCT 656
|
|
|
Db 19841 CGGGAGAGCAAGGCGACGACCTTGGGGGCGCTTCTTGAAGCCAGACAGATTTTCTCTC 19900
|
|
|
QY 657 GCGAGTCTGTGTCCTCCAC-----CAGTTTGTGTGTGCATCATCGAAAGAGGCTT 710
|
|
|
Db 19901 GCGAGTCTGTGTCCTCCACAGTTTGTGTCAGTTTGTGTGTGCATCATCAAAAGAGGCTT 19960
|
|
|
QY 711 GACCAATAAGAACTACCTCTAGCAGACCCCACTCCCGGTAAATATCCATGAAAGAGAA 770
|
|
|
Db 19961 GACCAATAAGAACTACCTCTAGCAGACCCCACTCCCGGTAAATATCCATGAAAGAGAA 20020
|
|
|
QY 771 CTCTGAGCTGCAAGAACTGTCAACCATCATGCAACCCGAGGGGAGCTTCTGAGC 830
|
|
|
Db 20021 CTCTGAGCTGCAAGAACTGTCAACCATCATGCAACCCGAGGGGAGCTTCTGAGC 20080
|
|
|
QY 831 ATGCGCATGATTTCTTGAATCATGCAAGAAAGGAGAGTAGAGCAAACTAGCCGAGA 890
|
|
|
Db 20081 ATGCGCATGATTTCTTGAATCATGCAAGAAAGGAGAGTAGAGCAAACTAGCCGAGA 20140
|
|
|
QY 891 GATTCCCTGAAAAATCTTGGGCAACAATGCTTGTGTGAAAGCTGATTTGAAAAAGAG 950
|
|
|
Db 20141 GATTCCCTGAAAAATCTTGGGCAACAATGCTTGTGT-----GAAAAAGAG 20188
|
|
|
QY 951 CAGAAATTTGAAAGAAATTTGAACATGAAACAGGAGCAAGATACATTCATCTTTGCA 1010
|
|
|
Db 20189 CAGAAATTTGAAAGAAATTTGAACATGAAACAGGAGCAAGATACATTCATCTTTGCA 20248
|
|
|
QY 1011 GGAATTGAGCATATACAAACCGGAAAGAACCATCATCTGTGAAGGCAAG-TTGAAGCT 1069
|
|
|
Db 20249 GGAATTGAGCATATATATACCAAGAAAGATCATCTGTGAAGGCAAGTTGAGGCT 20308
|
|
|
QY 1070 GTGGCAATGCTGAGATGAGATTTGAAGAAAGCTGTGAGGCTTTGAAATGATATGC 1129
|
|
|
Db 20309 GTGGCAATGCTGAGATGAGATTTGAAGAAAGCTGTGAGGCTTTGAAATGATATGT 20368
|
|
|
QY 1130 TGGCTGTTAACACCACTCCGATACTTCTCAGGCTGTACCCCATCACAGTTTGGC 1189
|
|
|
Db 20369 TGGCTGTTAATAGCACTCCGATACTTCTCAGGCTGTACCCCATCACAGGTTGGC 20428
|
|
|
QY 1190 CGTTCCGCAATCACTCTTATTCAGAGCAGAGATGTGAATCTTTCATCCCAACC 1249
|
|
|
Db 20429 CGTTCCGCAATCACTCTTATTCAGAGCAGAGATGTGAATCTTTCATCCCAACC 20488
|
|
|
QY 1250 AAGCTGTGGGCGCATATGCGGAAAGAGGGGCGACACATCAACAGCTGGCGAGATTG 1309
|
|
|
Db 20489 AAGCTGTGGGCGCATATATCAGAAAGAGGGGAGCACATCAACAGCTGGCGAGATTG 20548
|
|
|

QY 1310 CCGAGCCTTATCAAGATTGCCCCCTGGGAAAGGCCAGAGCTACGCCAAAGATGTCA 1369
|
|
|
Db 20549 CCAAGCCTCATCAAGATCGCTTGGGAAAGGCCAGAGCTCAACCAAAAGATGTCA 20608
|
|
|
QY 1370 TCATACCGGGGACCCGGAAGGCCAGTTCAAGGCCCCAGAGGAGCATCTTTGGAAATCGA 1429
|
|
|
Db 20609 TCATACCGGGGACCCGGAAGGCCAGTTCAAGGCCCCAGAGGAGCATCTTTGGAAATCGA 20668
|
|
|
QY 1430 AAGAGAAAATCTTTTAAACCCCAAGAAAGTGAAGCTGAGACGATATCAAGTGC 1489
|
|
|
Db 20669 AAGAGAAAATCTTTTAAACCCCAAGAAAGTGAAGCTGAGACGATATCAAGTGC 20728
|
|
|
QY 1490 CCTCTTCAACAGCTGCGGGGTGATTTGGCAAAAGTGGGCAAGACCTGTACACGACGA 1549
|
|
|
Db 20729 CCTCTTCAACAGCTGCGGGGTGATTTGGCAAAAGTGGGCAAGTGAATCTTGAATGAATCGACA 20788
|
|
|
QY 1550 ACTTAACAGTGCAGAAAGTCAATCGTCTCGTGAACAAAGCCAGATGAAATGAGGAAG 1609
|
|
|
Db 20789 ACTTAACAGTGCAGAAAGTCAATCGTCTCGTGAACAAAGCCAGATGAAATGAGGAAG 20848
|
|
|
QY 1610 TGATGTCAGAAATTATCGGGCACTTCTTTGTAGCCAGACTGCAACAGGCAAGTCAAGG 1669
|
|
|
Db 20849 TGATGTCAGAAATTATGGAACATCTTGTCTACCCCACTGCAATGACAAAGATCAAGG 20908
|
|
|
QY 1670 AAATTGTACACAGGTGAAGAGCAGAGAGCAAGAAATACCTCAGGAGTGGCTTCAAGC 1729
|
|
|
Db 20909 AAATTGTACACAGGTGAAGAGCAGAGAGCAAGAAATACCTCAGGAGTGGCTTCAAGC 20968
|
|
|
QY 1730 GCAGCAAGTGAAGCTCCCAACAGGCAACGCAAAACAAAGATGATGAGGCTTCCAG 1789
|
|
|
Db 20969 ACAGCAAGTGAAGCTCCCAACAGGCAACGCAAAACAAAGATGATGAGGCTTCCAG 21028
|
|
|
QY 1790 ACTGAAGATGAGACCAACGACGCCAGATCGGAGCAAAACCAAGACCATCTG 1849
|
|
|
Db 21029 ACTGAAGATGAGACCAACGACGCCAGATCGGAGCAAAACCAAGACCATCTG 21088
|
|
|
QY 1850 AGGAATGAGAAAGTGTGGGAGGCGGCGAGGAGCTCTCGAGGCGCTGAGAACCCAGGG 1909
|
|
|
Db 21089 AGGAATGAGAAAGTGTGGGAGGCGGCGAGGAGCTCTCGAGGCGCTGAGAAATCCAGGG 21148
|
|
|
QY 1910 GCCGAGGAGGGGCGGGAGAGGTGAGCCAGGTTTCCGAACCAACGAGGCC-CGCTCC 1968
|
|
|
Db 21149 GCCGAGGAGGGGCGGGAGAGGTGAGCCAGGTTTCCGAACCAACGAGGCCCGCTTCCAC 21208
|
|
|
QY 1969 GCCCCCAAGGCTTCTGCAAGCTTCAACCAATCACTTCAACCATCACTCGATCTCT 2028
|
|
|
Db 21209 CCCCATCAGGCTTCTGCAAGCTTCAACCAATCACTTCAACCATCACTCGATCTCTCT 21268
|
|
|
QY 2029 GAATCCCAAGGCTATCCCTTTTGTAGTGAATCAATAGTGAAGTGTTCAGAGCA 2088
|
|
|
Db 21269 TAACTCCCAAGGCTATCCCTTTTGTAGTGAATCAATAGTGAAGTGTTCAGAGCA 21326
|
|
|
QY 2089 AGCAAAATGACACCCCTTTTCTGTGCAATCGTCTCTGATCATGTGTATATATG 2148
|
|
|
Db 21327 AGCAAAATGACACCCCTTTTCTGTGCAATCGTCTCTGATCATGTGTATATATG 21386
|
|
|
QY 2149 AAAAGGAAAGTGTAGATATGTGGCCTGTGGTTACAGAGGCTCTGACGCGGTATA 2208
|
|
|
Db 21387 AAAAGGAAAGTGTAGATATGTGGCCTGTGGTTACAGAGGCTCTGACGCGGTATA 21446
|
|
|
QY 2209 TATTTTGAATATATATATCAATTAATCAATTAATCAATTAATCAATTAAT 2268
|
|
|
Db 21447 TATTTTGAATATATATATCAATTAATCAATTAATCAATTAATCAATTAAT 21501
|
|
|
QY 2269 TTTTCTTTCTTTTAAAGAAAGAGGCTTTCTGAGCTTTTAAAGAAATTA 2233
|
|
|
Db 21502 TTTTCTTTCTTTTAAAGAAAGAGGCTTTCTGAGCTTTTAAAGAAATTAAGTCTTTC 21561
|
|
|
QY 2324 TTTGGAGGTTCTACGCTGTAGAGAGGCTTTGAGGCCACCGCAAAATTCACCCAG 2383
|
|
|
Db 21562 TTTGGAGGTTCTACAGTATAGACAGAGGCTTTGAGGCCACCGCAAAATTCACCCAG 21621
|
|
|
QY 2384 AGGGAATCTGTGTGAGAGAGCACTCAGCGAGTTCTGATATCCTGTATGTCAACAG 2443
|
|
|

Db	21622	AGGGAATCTCAAGAGAAAGAAACAATCAGAGAGTTCTGGATCACTGTGATGTCAACAG	21681
QY	2444	AAGGATACCGTCTCCCTTGAAGAGAACTGTGTCACTCTCATGCTGTCTACCTATA	2503
Db	21682	AAGGATACCATCTCTTGAAGAGAAACTGTGTCACTCTCATGCTGTCTACCTATA	21741
QY	2504	CACCCATTCTCTTGGCTTCAAGAGTTTAACTGTGTTTTGCACTAGCTATATAT	2563
Db	21742	CACCCATTCTCTTGGCTTCAAGAGTTTAACTGTGTTTTGCACTAGCTATATAT	21801
QY	2564	CTGTGCTCTCTGTGTATATCTCCCTCCCTCCCTCCCTCCCTCTTCTCATCTCAT	2623
Db	21802	ATCATCTCTCTGTGTATATCTCCCTCCCTCCCTCCCTCCCTCTTCTCATCTCAT	21861
QY	2624	CTTTGAAATTCCTCATCCCTCCATCTCAATCCCGATCTACGACG--CCCCCCCCCA	2682
Db	21862	CTTTGAAATTCCTCATCCCTCCATCTCAATCCCGATCTACGACCTCATCTCCACA	21921
QY	2683	GGCAAGACAGTGTCTGAGTATCACTACACAAAGAAAGAAAGCAAGCAACAAAC	2742
Db	21922	GGGAAACAGTGTCTGAGTATCGATCAACAGAAAGAAAGAAAGTAAACACAAAC	21981
QY	2743	CAGCTCAACTTCACTTGGTCTACTCAAAAGAAACAAGTCAATGTACTTGTCTACG	2802
Db	21982	CAACTTAATTATACATTGGTACTCAAAAGAAACAAGTCAATGTACTTGTCTACG	22041
QY	2803	TTTTGAAAGAGAAAACAGAACCCACAAACCAACCAATCAACCAACAAAGAAAAAT	2862
Db	22042	TTTTGAAAGAGAAAACAGAACCCATCAACCAACCAATCAACCAACAAAGAAAAAT	22101
QY	2863	TCACAAATGAAAGATATATTTGTGCT--TTTGGCAATTTGGTGATTAAGCAATAT	2921
Db	22102	TCACAAATGAAAGATATATTTGTGCTTTTGGCAATTTGGTGATTAAGCAATATAT	22161
QY	2922	CAGAAAAATGATTCCTTTCTTAAAAAAAATGTGAGAGAAAGTAAATTTTCCAAAG	2981
Db	22162	CAGAAAAATGATTTCTTTCTTT--AAAAAAATGTGAGAGAAAGTAAATTTTCCAAAG	22218
QY	2982	GTTGTGGCCACAGGCGTTAAATTCACAGATTTTAAACGAAAAACACACAGAGAA	3041
Db	22219	GTTGTGGCCACAGGCGATTAATTTACAGATTTTAAATGGGAAAAACACACAGAAAA	22278
QY	3042	GCTACTCTAGGTG--TTTTTACCTCAGACCTTGTCTGTGTTTCCCTTAGAGATTTGT	3100
Db	22279	GCTCTCTCAGGTGTTTTTTACCTCAGACCTTGTCTGTGTTTCCCTTAGAGATTTGT	22338
QY	3101	AAAGCTGATGTGAGACATTTTTTTT--TTTTTAAATAAAATGAGTGAAGAAAAAT	3159
Db	22339	AAAACTGATGTGAGACATTTTTTTTAAATTTTTTAAATAAAATGAGTGAAGAAAAAT	22398
QY	3160	AAAGATATCAACTGACCTGAGAGAGTGAACAGTCCAAATGTGCAACAGCTTTCTGAA	3219
Db	22399	AAAGATATCAACTGACAGCTGAGAGAGTGAACAGTCCAAATGTGCAACAGCTTTCTGAA	22458
QY	3220	TTGTCTTCGCTAGCCAAAGAACNAATATGCGCTTTCTTTGACAAACTTTGAAATGTT	3279
Db	22459	TTGTCTTCGCTAGCCAAAGAACNAATATGCGCTTTCTTTGACAAACTTTGAAATGTT	22518
QY	3280	ATTT 3283	
Db	22519	ATTT 22522	
RESULT 12			
LOCUS	AC104980	187226 bp	DNA linear PRI 31-MAR-2002
DEFINITION	Homo sapiens chromosome 8, clone RP11-281D17, complete sequence.		
ACCESSION	AC104980		
VERSION	AC104980.5	GI:19849375	
KEYWORDS	HTG.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

1 Ekaryota; Metazoa; Chordata; Catarrhini; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
2 (Pages 1 to 187226)
3 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
4 Homo sapiens chromosome 8, clone RP11-281D17
5 Unpublished
6 2 (Pages 1 to 187226)
7 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
8 Anderson, S., Barina, N., Baeten, V., Boguslavsky, L., Bouthgalter, B.,
9 Brown, A., Camarata, J., Campiolo, A., Chang, J., Chazaro, B.,
10 Choepel, Y., Colangelo, M., Coppolino, S., Collins, A., Collymore, A., Cook, A.,
11 Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
12 Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J., Gage, S.,
13 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
14 Huges, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
15 Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K.,
16 Lamasarres, R., Landers, T., Lehotzky, J., Levine, R., Liu, G.,
17 Maclean, C., MacDonald, P., Mayo, J., Marquis, N., Matthews, C.,
18 McCarthy, M., McEwan, P., McKernan, K., McNeheaters, R., Meldrum, J.,
19 Menelus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
20 Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
21 Oliver, J., Peterson, C., Phunhkan, P., Pierre, N., Pollara, V.,
22 Raymond, C., Retta, R., Ryback, M., Riley, R., Rise, C., Rogov, P.,
23 Roman, J., Rossetti, M., Roy, A., Santos, R., Schauer, S., Stojanovic, N.,
24 Seaman, S., Severly, P., Spencer, B., Strange-Thomann, N.,
25 Straus, K., Subramanian, A., Talamas, J., Teeffaye, S., Theodore, J.,
26 Topham, K., Travins, M., Trivis, L., Triggillo, V., Vassiliou, H.,
27 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.-Y., Young, G.,
28 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

1 TITLE
2 Direct Submission
3 JOURNAL
4 Submitted (22-DEC-2001) Whitehead Institute/MIT Center for Genome
5 Research, 320 Charles Street, Cambridge, MA 02141, USA
6 3 (bases 1 to 187226)
7 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
8

REFERENCE	JOURNAL	TITLE
AUTHORS		
Anderson, S., Barna, N., Baetien, V., Blodum, L., Boguslavsky, J., Bouhgalter, B., Brown, A., Camarata, J., Campiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamares, R., Lander, E., Lehocsky, J., Levine, R., Linblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L., Mihova, T., Meng, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivier, J., Peterson, C., Phunkharn, P., Pierrat, N., Pollara, V., Raymond, C., Retta, R., Ribick, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramaniam, A., Talams, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.U., Young, G., Zainoun, J., Zembek, L., Zimmer, L. and Zody, M.	Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA 4 (bases 1 to 187226)	

TITLE
JOURNAL

COMMENT

Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Riese, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Struss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Trivis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, W.

Direct Submission
Submitted (31-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 31, 2002 this sequence version replaced gi:19683374.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: 122878
Center clone name: 281_D_17

FEATURES
source

Location/Qualifiers
1..187226
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-281D17"
/clone_lib="RPCT-11 Human Male BAC"
1..5915
/rpt_family="L1PA5"
5916..5943
/rpt_family="AT_rich"
5944..6477
/rpt_family="L1"
8563..8886
/rpt_family="MER2"
repeated_region
complement(9296..9555)
/rpt_family="L2"
repeated_region
complement(9865..10094)
/rpt_family="MIR"
repeated_region
complement(10346..10504)
/rpt_family="AluSg/x"
repeated_region
complement(10821..11111)
/rpt_family="AluSc"
11162..11300
/rpt_family="MER47A"
11301..11596
/rpt_family="AluSx"
11597..11836
/rpt_family="MER47A"
12185..12319
/rpt_family="FLAM_C"
12520..12822
/rpt_family="AluSp"
repeated_region
complement(12885..13167)
/rpt_family="AluYb"
repeated_region
complement(13230..13379)
/rpt_family="MIR"
repeated_region
complement(14179..14479)
/rpt_family="AluSg"
repeated_region
complement(14630..14747)
/rpt_family="MIR"
14748..15055
/rpt_family="AluSq"
repeated_region
complement(15056..15145)
/rpt_family="MIR"
repeated_region
complement(15206..15400)
/rpt_family="MIR"

Query Match	Best Local Similarity	89.4%	Pred. No. 0;	Mismatches 211;	Indels 113;	Gaps 16;
247	CGTAAAGTGAATTCATGAGGAAATATCATGGAATGATTAATCAATCTCTAAAGGCTA	306				
Db	79640	CGTAAAGTGAATTCATGAGGAAATATCATGGAATGATTAATCAATCTCTAAAGGCTA	79699			
QY	307	AGAGACGAGAAATTCAGATTCGAAACATCCTCTCATCCTGAGTGGAGTGTGGAT	366			
Db	79700	AGAGACGAGAAATTCAGATTCGAAACATCCTCTCATCCTGAGTGGAGTGTGGAT	79759			
QY	367	GGACTTTGGCTCAATATGAGACAGTGGAGAAATGTGGAACAAGTCAACAGACAGAA	426			
Db	79760	GGACTTTGGCTCAATATGAGACAGTGGAGAAATGTGGAACAAGTCAACAGACAGAA	79819			
QY	427	ACGCGCGTGTCAACGCAATATGCAACAAGAGAGAGAGAAATAGCATGAGAGAG	486			
Db	79820	ACGCGCGTGTGTCAATATGCAACAAGAGAGAGAGAGAAATAGCATGAGAGAG	79878			
QY	487	CTAAGCGGAGCATGATTGAGAACTACTCTTCAAGATTTCTCAATCCCGATGAGAG	546			
Db	79879	CTAAGCGGAGCATGATTGAGAACTACTCTTCAAGATTTCTCAATCCCGATGAGAG	79938			
QY	547	GTAGGCT-CCCTTCCGCCCTCAGGAGCCCAAGCGTGGGAGCACTTTCCCGGAGCA	605			

Db	79939	GTGAGCTCCCCCTTTACCCCTCAGGAGCCAGGGGTGGAGCACTCTT - CCGGAGCA	79997
Qy	606	AGGCGACGCCCCCTGGGGGCACTTCTCAGGCGCAACAGATTGATTTCCGCTCGGATCCT	665
Db	79998	AGGTCACG - CCTGGGGGCTCTTCTCAGGCGGAGACAGACTGATTTCCGCTCGGATCCA	80056
Qy	666	GGTCCCAACCCAGTTTGTGTGTGCATCATCGGAAAGAGGGCTTGAACAATAAGAACAT	725
Db	80057	----- ACCAGTTTGTGTGTGCATCATCGGAAAGAGGGCTTGAACAATGTGAACAT	80109
Qy	726	CACTTAAGCACCCAGTCCCGGGTAGTATCCATAGAAAAGAACCTCGAGCTCAGAC	785
Db	80110	CACTAGACAGACCCAGTCCCGGGTAGCATCATAGAAAAGAACCTCGAGCTCAGAC	80169
Qy	786	GAACTGTGCAACATCCATGCCACCCAGAGGGGACTTTCGAGCATGCCGATGTTCT	845
Db	80170	GAACTGTGCAACATCCATGCCACCCAGAGGGGACTTTCGAGCATGCCGATGTTCT	80229
Qy	846	TGAATCATGCAAGAAAGAGGCAATGAGACCAACTAGCCGAAGATTTCTTGAAAT	905
Db	80230	TGAATCATGCAAGAAAGAGGCAATGAGACCAACTAGCCGAAGATTTCTTGAAAT	80288
Qy	906	CTTGGCACACATGGCTTTGTGTGTAACATGATTGAAAAGAGGCAAGAAATTTGAAGA	965
Db	80290	CGTGGCCACATGGCTTTGTGTGTAACATGATTGAAAAGAGGCAAGAAATTTGAAGA	80349
Qy	966	AATTGAACATGAACAGGAGCAACAGATTAACATCTCTTTGCAGATTTGACATATA	1025
Db	80350	AATTGAACATGAACAGGAGCAACAGATTAACATCTCTTTGCAGATTTGACATATA	80408
Qy	1026	CAACCCGAAAGAACCATCATCTGTGAAGGCAAGTTGAGGCTGTGCAGTGTGAGAT	1085
Db	80410	CAACCCGAAAGAACCATCATCTGTGAAGGCAAGTTGAGGCTGTGCAGTGTGAGAT	80465
Qy	1086	AGAGTTTATGAAGAAAGCTGCTGAGGCTTTGAAAATGATAGTGTGCTGTAAACCCA	1145
Db	80470	AGAGTTTATGAAGAAAGCTGCTGAGGCTTTGAAAATGATAGTGTGCTGTAAACCCA	80522
Qy	1146	CTCGGATATG - TTCTCCAGCTGTACCCCATCAACAGTTTGGCCCGTTCCCGCATATC	1204
Db	80530	CTCGGATATGCTTCTCCAGCTGTACCCCATCAACAGTTTGGCCCGTTCCCGCATATC	80589
Qy	1205	ACTCTTATCCAGAGCAGAGATTGTGAATCTTTCATCCCAACCCAGGCTGTGGGCGCA	1264
Db	80590	ACTCTTATCCAGAGCAGAGATTGTGAATCTTTCATCCCAACCCAGGCTGTGGGCGCA	80644
Qy	1265	TCATGGGAAAGAGGGGGCACATCAAAACAGCTGGGAGATTGGCCGGAGCCCTATCA	1324
Db	80650	TCATGGGAAAGAGGGGGCACATCAAAACAGCTGGGAGATTGGCCGGAGCCCTCA ---	80700
Qy	1325	AGATTGCCCTCGCGGAAGCCCAAGCGTCAGCGAAAGGATGTATCATCACCGGCGCAC	1384
Db	80707	-----	80700
Qy	1385	CGGAAGCCCATGTTCAAGGCCAGGAGCGGATCTTTGGGAAACTGAAAAGGAAAACTTCT	1444
Db	80707	----- TCAAGGCCAGGAGCGGATCTTTGGGAAACTGAAAAGGAAAACTTCT	80705
Qy	1445	TTTAAACCCAAAGAAAGTGAAGCTGTGAAGGGCAGATTCAGAGTCCCTTCCACAGCTG	1504
Db	80755	TTTAAACCCAAAGAAAGTGAAGCTGTGAAGGGCAGATTCAGAGTCCCTTCCACAGCTG	8081
Qy	1505	GCCGGGTGTTGGCAAAGGTGGCAAGACGCTGAACGAATCTCAGAACTTAAACCGCTGAG	1564
Db	80815	GCCGGGTGTTGGCAAAGGTGGCAAGACGCTGAACGAATCTCAGAACTTAAACCGCTGAG	8087
Qy	1565	AAGTCATCGTGCCTGTGACCAACGCCAGATGAATGAGAAATGAGAAATGTGTCGAATTA	1624
Db	80875	AATCATTTGTGCTGTGACCAACGCCAGATGAATGAGAAATGAGAAATGTGTCGAATTA	80933
Qy	1625	TCGGGCACTTCTTTGCTAGCCGACTGCAACGCCGAATCAAGGAAATTTGTAAACAGG	1684

Db	80935	TCGGGCACTTCTTTGCTAGCCAGACTGCAACAGCGCAGATATCAAGGAAACTGTACAAAGG	80994
QY	1685	TGAAGCAGCAGGAGCAAGAAATACCTCAGGAGTGCCTCAACGCGCAGCACTGAGC-C	1743
Db	80995	TGAACACAGAGGATGAAGAAATACCTCAGGAGTGCCTCAACGCGCAGCACTGAGGCC	81054
QY	1744	TCCCACAGGCACACGAAAAACAACGATGAATGTAGCCCTTCCAAACCTGACAGATGA	1803
Db	81055	TCCCACAGGCACACGAAAAACAACGATGAATGTAGCCCTTCCAAACCTGACAGAAATGA	81114
QY	1804	GACCAAAACGACCCAGCCAGATTCGGGAGCAAAACAAGAACCATCTGAGATGAAGATC	1863
Db	81115	GACCAAAACGACCCAGCCAGATTCGGAGCAAAACAAGAACCATCTGAGATGAAGATC	81174
QY	1864	TGCGAGGCGGCGCAAGGACTGTCCCGAGGCCCTGAGAACCCCAAGGCGCAGAGGCGG	1923
Db	81175	TGCGAGGCGGCGCAAGGACTGTGTACAGAGGCCCTGAGAAATCCAGGGCGCAAGGAGGCG-G	81233
QY	1924	GGAAGAGTCAAGCAGGTTTGGCCAGAACCAACGAGCCCGCGCTCCGCGCCCAAGGGCTTC	1983
Db	81234	GGAAGAGTCAAGCAGGTTTGGCCAGAACCTCCGGGCGCTGCTCTCACTCCACAGACTTC	81293
QY	1984	TGCAGGCTTCAGGCATCGACTTCAACCATCACTCGGATCTCTCTGAATCCACGACGC	2043
Db	81294	TGCA- GCTTCAGCAATCCACTTCACATTCACATCTCAATCTCTCTTAACCTCCCAATGC	81352
QY	2044	TATCCCTTTTAACTGTAACATAGTGAACGTGTTCAAAGCCAAAGCAAAATGCAACC	2103
Db	81353	TGTACTTTTGTGTGAAGTACATAGTGAACGTGTTCAAAGCCAAAGCAAAATGCAACC	81412
QY	2104	CTTTTCTGTGGCAATCGCTCTGTACATGTGTGTACATATGAAAGGGAATGTGA	2163
Db	81413	CTTTTCTGTGGCAAA----TCTGTACATGTGTGTGCAATATGAAAGGGAATGTGA	81468
QY	2164	AGATATGTGGCCCTGTGGGTTACACAGGTCCTGACAGCGGTAATATATTTAGAAATAT	2223
Db	81469	AGATATGTGGCCCTGTGGGTTACACAGGTCCTGACAGCGGTAATATGTTTAGAAATAC	81528
QY	2224	AATCAAAATACTCAACATACCTCAATTTTAAATCAATTAATTTTCTTTTCTTTTA	2283
Db	81529	AATCAAAATACTCAACATATCTCAATTTTAAATCAATTAATTTTCTTTTCTTTTAA	81588
QY	2284	AAGAGAAAGCAGGCTTTTCTAGCTTTAAAGAAATAAGCTTTGGGAGTCTACCGGT	2343
Db	81589	AAGAGAAAGCAGGCTTTTCTAGCTTTAAAGCTCT---TCTTTGGGAGTCTACCGGT	81644
QY	2344	AGAGAGAGCTTTGAGGCGCACCGCACAAAATTCACCCAGAGGAAATCTGTCGGAAG	2403
Db	81645	AGAGAGAGCTTTGAGGTCACCTACACAAAAATTCACCCAGAGGAAATCTTTCGGAAG	81704
QY	2404	ACACTCAGCGCACTTCGATCACTCTGTATGTCAACAGAGGATACCGTCTCTTGA	2463
Db	81705	ACACTCAGCGCACTTCGATCACTCTGTATATCAACAGAGGATACCAATCTCTTGA	81764
QY	2464	AGAGAAACTCTGTCACTCTCATGCTGTCTACTCATACACCATTTCTTTGGCTC	2523
Db	81765	AAAGGAAACTCTGTCACTCTCATGCTGTCTACTCATACATCATTTCTTTGGCTC	81824
QY	2524	ACAGGTTTAACTGTTTTTGGATCTGTATATTAATCTC--TGTCTCTCTGTTT	2581
Db	81825	ATGGGTTTAACTGTTTTTGGATCTGTATATTAATCTCTATCTCTCTCTGTTT	81884
QY	2582	ATGCTCCCTCCCTCCCTCCCTCTTCTTCATCTCCATCTTTTGAATTTCTCATC	2641
Db	81885	ATGCTCCCTCCCTCCCTCCCTCTTCTTCATCTCCATCTTTTGAATTTCTCATC	81944
QY	2642	CTTCATCTCAATCCGTAATTAAGCA-----CCCGCCCCCGCCAGGAAAGCACT	2693
Db	81945	CTTCATCTCAATCCCAATATCTAGCAACACACCTTGCCCCCGCCAGGAAAGCACT	82004
QY	2694	GCTCTGAGTATCATATACACAAAGGAAACAAAGCGAAACACAAACCACTCACT	2753
Db	82005	GCTTTGAGTATCCCAACACAAAGGAAACAAAGCGAAACACACACGCTCACTCACT	8206

QY 966 AATTGAACATGAAAGAGGACCAAGATAACATCTCATCTTTGACAGATTGAGCATATA 1025
Db 93637 AATTGAACATGAAAGAGGACCAAGATAACATCTCATCTTTGACAGATTGAGCATATA 93578
QY 1026 CAACCCGAAAGAACATCATCTGTGAAAGGCAAGTTGAGGCTGTGCCAGTGTGAGAT 1085
Db 93577 CAACCCGAAAGAACATCATCTGTGAAAGGCAAGTTGAGGCTGTGCCAGTGTGAGAT 93518
QY 1086 AGAGATTATGAAAGAGCTGCGTGAAGGCTTTGAAAAATATATGCTGGCTTTTAAACCCA 1145
Db 93517 GAGAGATTATGAAAGAGCTGCGTGAAGGCTTTGAAAAATATATGCTGGCTTTTAAACCCA 93458
QY 1146 CTCCGATAC-TTCCAGACCTGTACCCCATCACAGTTGGCCCGTCCCGCATCATC 1204
Db 93457 CTCCGATACCTTTCTCCAGCTGTACCCCATCACAGTTGGCCCGTCCCGCATCATC 93398
QY 1205 ACTTTATCCAGAGAGAGATTGTGAATCTTTTCATCCCAACCCAGGCTGTGGCGCA 1264
Db 93397 ATTCTTATCAAGAGAGAGATTGTGAATCTTTTCATCCCAACCCAGGCTGTGGGTACCA 93338
QY 1265 TCATCGGGAAGAAAGGGGCAACATCTCAACAGCTGGCGAGATTGCCGAGCTCTTATCA 1324
Db 93337 TCATCGGGAAGAAAGGGGCAACATCTCAACAGCTGGCGAGACTGGCCAGAGCTTCCA--- 93281
QY 1325 AGATTGCCCTGCGGAAAGCCAGACGTCAAGGAAAGATGATCATCATCCGGGCGAC 1384
Db 93280 ----- 93281
QY 1385 CGAAGCCCAAGTTCAAGGCCCAAGAGCCGATCTTTGGGAACTGAAAGAGAAAACTTCT 1444
Db 93280 -----TCAAGGCCCAAGAGCCGATCTTTGGGAACTGAAAGAGAAAACTTCT 93233
QY 1445 TTAACCCCAAGAAAGAGTGAAGCTGGAAGGCAATATAGTGCCTCTTCCACAGCTG 1504
Db 93232 TTAACCCCAAGAAAGAGTGAAGCTGGAAGGCAATATAGTGCCTCTTCCACAGCTG 93173
QY 1505 GCCCGGATGATGCAAAAGTGGCAAGACCTGAAAGCAATGAGAACTTAAACAGTGAAG 1564
Db 93172 GCCCGGATGATGCAAAAGTGGCAAGACCTGAAAGCAATGAGAACTTAAACAGTGAAG 93113
QY 1565 AAGTATCTGTGCTCTGTGACCAAAAGCCAGATGAATAATGAGAAAGTATCTCAGAAATTA 1624
Db 93112 AAGTATCTGTGCTCTGTGACCAAAAGCCAGATGAATAATGAGAAAGTATCTCAGAAATTA 93053
QY 1625 TCGGGGCACTCTTTGTCAGCCAGCTGACAGCGCCAGATATAGGAAATTTGTAACAAGG 1684
Db 93052 TCGGGGCACTCTTTGTCAGCCAGCTGACAGCGCCAGATATAGGAAATTTGTAACAAGG 92993
QY 1685 TGAAGCAGCAGAGCAAAATTAACCTCAGGAGTGCCTCAACAGCGAGCAAGTGAGG-C 1743
Db 92992 TGAAGCAGCAGAGCAAAATTAACCTCAGGAGTGCCTCAACAGCGAGCAAGTGAGG-C 92933
QY 1744 TCCCAACAGGCAACAGCAAAACAAAGATGAATGATGAGCTTCCAAACCTGACAGAAATGA 1803
Db 92932 TCCCAACAGGCAACAGCAAAACAAAGATGAATGATGAGCTTCCAAACCTGACAGAAATGA 92873
QY 1804 GACCAACAGCAGCAGCCAGATCGGGAGCAAAACCAAGACCATTTGGAATGAATGATC 1863
Db 92872 GACCAACAGCAGCAGCCAGATCGGGAGCAAAACCAAGACCATTTGGAATGAATGATC 92813
QY 1864 TGGGAGGCGGCAAGGAGCTGTGCGAGGCTTGAGAAACCCAGGGGCGGAGGAGGCG 1923
Db 92812 TGGGAGGCGGCAAGGAGCTGTGCGAGGCTTGAGAAATCCAGGGGCGGAGGAGGCG-G 92754
QY 1924 GGGAGGCTCAGCAGGTTTGCAGAAACACAGGAGCCCGCTCCCGCCCGCAGAGGCTTC 1983
Db 92753 GGGAGGCTCAGCAGGTTTGCAGAAACACAGGAGCCCGCTCCCGCAGAGGAGCTTC 92694
QY 1984 TGGAGGCTCAGCAGCATTCACTTCAACCATCACTCGATCTCTCTGAACTTCCACAGAGC 2043
Db 92693 TGGAGGCTCAGCAGCATTCACTTCAACCATCACTCGATCTCTCTGAACTTCCACAGAGC 92635

QY 2044 TATCCCTTTAGTTGAACATAATGAGTGAAGCTGTCAAGCCAGCAAAATGCAAC 2103
Db 92634 TATCCCTTTAGTTGAAGTGAAGTGAAGCTGTCAAGCCAGCAAAATGCAAC 92575
QY 2104 CTTTTCCTGTGGCAAAAGCTCTGTGATCATGTGTATCATATTAAGAAAGGAAAGTTTA 2163
Db 92574 CTTTTCCTGTGGCAAAAGCTCTGTGATCATGTGTATCATATTAAGAAAGGAAAGTTTA 92519
QY 2164 AGATATGTGCTGTGGTTTACACAGAGGTGCTGACAGGGAATATATTTTGAATAAT 2223
Db 92518 AGATATGTGCTGTGGTTTACACAGAGGTGCTGACAGGGAATATATTTTGAATAAT 92459
QY 2224 ATATCAATTAACCTCACTAATCTCAATTTTATATATTAATTTTTCCTTTT 2283
Db 92458 ATATCAATTAACCTCACTAATCTCAATTTTATATCAATTAATTTTTCCTTTT 92399
QY 2284 AAGAGAAAGCAGGCTTTCTAGACTTTTAAAGATTAAGCTTTGGGAGGTCTCAGGCT 2343
Db 92398 AAGAGAAAGCAGGCTTTCTAGACTTTTAAAGATTTAAAGCT---TTCTTGGAGGTTACAGGCT 92343
QY 2344 AAGAGAGCTTTGAGGCAACCCGACAAATTAACCCAGAGGAAATCTGTCGGAAGG 2403
Db 92342 AAGAGAGCTTTGAGGCAACCCGACAAATTAACCCAGAGGAAATCTGTCGGAAGG 92283
QY 2404 AACCTCAGGCAATTCGATACCTGTGTATGTCAACAGAAAGGATACGCTCTCTTGA 2463
Db 92282 AACCTCAGGCAATTCGATACCTGTGTATGTCAACAGAAAGGATACGATCTCTTGA 92223
QY 2464 AAGAGAACTGTGACCTCTCATGCTGTAGCTGTATACCATTTCTTTGCTC 2523
Db 92222 AAGAGAACTGTGACCTCTCATGCTGTAGCTGTATATATTTCTCATCTCTCTGTT 92163
QY 2524 ACAGTTTAAACTGTGTTTTTGTGCAACTGTATATAATTC--TGTCTCTCTGTT 2581
Db 92162 ATGGGTTTAAACTGTGTTTTTGTGCAACTGTATATAATTTCTCATCTCTCTGTT 92103
QY 2582 ATCTCTCCCTCCCTCCCTCCCTCTCTCTGCAATCTCATTTCTTTGAATTTCTCTATC 2641
Db 92102 ATCTCTCCCTCCCTCCCTCCCTCTCTCTGCAATCTCATTTCTTTGAATTTCTCTATC 92043
QY 2642 CCTCATCTCATCTCCGATCTAGCA-----CCCCCCCCCCCCAGGCAAGCAAGT 2693
Db 92042 CCTCATCTCATCTCCGATCTAGCAACACACACACACACACACCTGCCCCCAGGAAAGCAAGT 91983
QY 2694 GCTGTGATATCATCATCAACAAAGGAAACAAAGCAACACACACACACAGCTCAACT 2753
Db 91982 GCTGTGATATCCCAACACAAAGGAAACAAAGCAACACACACACAGCTCAACT 91923
QY 2754 TACACTTGGTATCAATCAACAAAGGCAAGTCAATGATCTTGTCTTGGAGGAG 2813
Db 91922 TACACTTGGTATCAATCAACAAAGGCAAGTCAATGATCTTGTCTTGGAGGAG 91863
QY 2814 GAAACAGAGAACCAACCAACCAATCAATCAACCAACCAAGAAATTTCCAGATGAA 2873
Db 91862 GAAACAGAGAACCAACCAACCAATCAATCAACCAACCAAGAAATTTCCAGATGAA 91803
QY 2874 AGAATGATTTGTCTTTTGAATTTGTATTAAGCAATCAATTTACAGAAATGAT 2933
Db 91802 AGAATGATTTGTCTTTTGAATTTGTATTAAGCAATCAATTTACAGAAATGAT 91743
QY 2934 TCCCTTCTTAAAAAAAATGTGAGAGAAAGTGAATTTTACAGAGTTGTGGCCA 2993
Db 91742 TCCCTTCTTAAAAAAAATGTGAGAGAAAGTGAATTTTACAGAGTTGTGGCCA 91683
QY 2994 GGGGTTTAAATTCAGATTTTAAACGAGAAACACACAGAGAGCTACTCAGGT 3053
Db 91682 GGGGTTTAAATTCAGATTTTAAATGAGAAACACACAGAGAGCTACTCAGGT 91623
QY 3054 G-TTTTAACTCAGCACTTGTCTTGTGTTTCTTTAGAGATTTGTAAGCTGATAGT 3112
Db 91622 GTTTTAACTCAGCACTTGTCTTGTGTTTCTTTAGAGATTTGTAAGCTGATAGT 91563
QY 3113 TGGAGCAATTTTATTTTAAATGAATGAGTTGAGAAAAAATGAATATCAACTG 3172

Db 91562 TGGAGCA-----TTTATTTTCTTAATAAATGAGTTGAAAAAATAGATATACAA 91507
QY 3173 CCAGCTGAGAGAGGTGACAGTCCAGTGTGCAACAGCTGTTTGAATTTCTTCGCTA 3232
Db 91506 CCAAGCTGAGAGAGGTGAGAGTCCATTTGTGCAACACTGTTCTGAATTTCTTCGCTA 91447
QY 3233 GCCAAGAACCMATATGCGCTTCTTTTGGACCAACCTTGAATTTCTT 3282
Db 91446 GCCAAGAACCTATATGCGCTTCTTTTGAACCAACCTTGAATTTAT 91397

RESULT 14
BC023758
LOCUS
DEFINITION Mus musculus RIKEN cDNA C330012H03 gene, mRNA (cDNA clone
IMAGE:5354659), partial cds.
ACCESSION BC023758
VERSION
KEYWORDS BC023758.1 GI:23958572
SOURCE
ORGANISM Mus musculus (house mouse)
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3557)
Strausberg,R.L., Pelngold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shermen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Scheetle,M., Soares,M.B., Bonaldo,M.P., Casavant,T.L.,
Schertz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carinci,P., Frange,C., Raha,S.S., Loguclano,N.A., Peters,G.J.,
Abramson,R.D., Mullany,S.J., Bosak,S.A., McKean,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,B.J., Lu,X., Gibbs,R.A.,
Fahy,U., Helton,E., Kettelman,M., Madan,A.C., Rodighiero,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Boutard,G.G., Blakesley,R.W., Touchman,D.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smallus,D.B.,
Scherer,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
JOURNAL
MEDLINE
PUBMED 12477932
REFERENCE
2 (bases 1 to 3557)
Strausberg,R.
Direct Submission
Submitted (05-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIG-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgaps-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Contact: nisc.mgc@hgti.nih.gov
Web site: http://www.nisc.nih.gov/
Ahner,N., Ayala,K., Beckstrom-Steenberg,S.M., Benjamin,B.,
Blakesley,R.W., Boufard,G.G., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-U., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
Maduro,Q.L., Masello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantirpop,S., Thomas,P.J., Touchman,J.W.,
Turgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,

Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
Series: IRAC Plate: 55 Row: 0 Column: 24
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Genomescan gene prediction.
location/Qualifiers
1..3557
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5354659"
/tissue_type="Mammary tumor, C3(1)-Tag model, infiltrating
ductal carcinoma, 5 month old virgin mouse."
/clone_lib="NCI_CGAP_Mam6"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

ORIGIN
Query Match 58.6%; Score 1924.2; DB 10; Length 3557;
Best Local Similarity 79.7%; Pred. No. 0; Mismatches 459; Indels 229; Gaps 28;
Matches 2708; Conservative 0;
58 GGAAGAGACGAGATATGAAACAAGCTTTACATCGGAAACCTGAGCCCGCGTACCGCC 117
23 GGAAGAGACCGAGATATGAAACAAGCTTTACATCGGAAACCTGAGCCCGCGTACCGCC 82
QY 118 GAGCACTTCGCGACGCTCTTTGGGAGACAGAAAGCTCCCTGCGGAGCAGGTCTCTG 177
Db 83 GAGCACTTCGCGACGCTCTTCGCGGAGACAGAAAGCTCCCTGCGGAGAGGTCTTACTC 142
QY 178 AAGTCGCGTACGCGCTTGTGACTACCCGACCAAGATGCGGCATCGCCCATCGAG 237
Db 143 AAGTCGCGTACGCGCTTGTGACTACCCGACCAAGATGCGGCATCGCCCATCGAG 202
QY 238 ACCCTCTCGGTTAAAGTGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATG 297
Db 203 ACCCTCTCGGTTAAAGTGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATG 262
QY 298 AAAAAGCTAAGAGACGAGAAATTCAGATTGCAAAATTCCTTCACCTGCACTGAGAG 357
Db 263 AAAAAGCTAAGAGACGAGAAATTCAGATTGCAAAATTCCTTCACCTGCACTGAGAG 322
QY 358 GTTGGATGGAATTTTGGCTCAATATGGAAGAGAGAGATGGAATGGAATGGAATGGA 417
Db 323 GTTGGATGGAATTTTGGCTCAATATGGAAGAGAGAGATGGAATGGAATGGAATGGA 382
QY 418 GACACAGAAACCGCCGTTGTCAACGTACATATGCAACAGAGAGAGAGAGAGAGAGAG 477
Db 383 GATACAGAAACCGCCGTTGTCAACGTACATATGCAACAGAGAGAGAGAGAGAGAGAG 442
QY 478 ATGGAAGAGTAAAGCGGAGATCAGTTTGAAGACTACTCCTCAAGATTTCTCAATCCG 537
Db 443 ATGGAAGAGTAAAGCGGAGATCAGTTTGAAGACTACTCCTCAAGATTTCTCAATCCG 502
QY 538 GATGAAGAGTAAAGCGGAGATCAGTTTGAAGACTACTCCTCAAGATTTCTCAATCCG 597
Db 503 GATGAAGAGTAAAGCGGAGATCAGTTTGAAGACTACTCCTCAAGATTTCTCAATCCG 541
QY 598 CGGAGACAG 657
Db 542 CGGAGACAG 601
QY 658 CGGATCTGATCCCAACCGAGTTTGTGAGCATATGGAAGAGAGAGAGAGAGAGAGAG 717
Db 602 CGGATCTGATCCCAACCGAGTTTGTGAGCATATGGAAGAGAGAGAGAGAGAGAGAG 661
QY 718 AAGAACTACTAG 777
Db 662 AAGAACTACTAG 721

QY	778	GCTGCAGAGAAGCCTGTCAACCATCAATGCCACCAGAGGGAGCTTTCTAAGCATGCGC	837
Db	722	GCTGCAGAGAAGCCTGTCAACCATCAATGCCACCAGAGGGAGCTTTCTAAGCATGCGC	781
QY	838	ATGATTTCTTGAAATTCATGCAAGGAAGGCAAGATGAAACCAACTAGCCGAAAGATTCTT	897
Db	782	ATGATTTCTTGAAATTCATGCAAGGAAGGCAAGATGAAACCAACTAGCCGAAAGATTCTT	841
QY	898	CTGAAAAATCTTGGCAACAATGGCTGTTGGTGAAGACTGATTGAAAAAGAGCGAAGAT	957
Db	842	CTGAAAAATCTTGGCAACAATGGCTGTTGGTGAAGACTGATTGAAAAAGAGCGAAGAT	901
QY	958	TTGAAAGAAATTTGAAATGTAAGAACAGGACCAAGATTAACATCTCATCTTTGCAAGATTGG	1017
Db	902	CTGAAAGAAATTTGAAATGTAAGAACAGGACCAAGATTAACATCTCATCTTTGCAAGATTGG	961
QY	1018	AGCATATTAACACCCGGAAGAAACCATCACTGTGAAGGGGACAGTTGAGGCTGTGCCAGT	1077
Db	962	AGCATATTAACACCCGGAAGAAACCATCACTGTGAAGGGGACAGTTGAGGCTGTGCCAGT	1021
QY	1078	GCTGAGATGAGATTATGAAAGAGCTGCGTGAAGGCTTTGAAAAATGATGTGCTGCTTT	1133
Db	1022	GCTGAGATGAGATTATGAAAGAGCTGCGTGAAGGCTTTGAAAAATGATGTGCTGCTTT	1081
QY	1138	AAC-----	1140
Db	1082	AACCAACAAGCCAAATCTGATCCAGGGCTTAAACCTGAGTGAATTGGCATCTTTTGCAGCT	1141
QY	1141	-----	1140
Db	1142	GGACTGTGTGCTTCTCTCCACACAGAGGGCCCGGTGAAGTTCCCCCAAGTCTCCAT	1201
QY	1141	-----ACCACTCGGGATTACTTTCTCAGCTGTGAACCCCATCAACAGTTTGGC	1188
Db	1202	CACCCCTTTGCTAACCCACCTCCGGATTACTTCTCAGCTGTGAACCCCATCAACAGTTTGGC	1261
QY	1189	CCGTTCCCGCATCATCACTCTTAATCCAGAGAGAGATTGGAATCTCTTATCCCAAC	1248
Db	1262	CCATTTCCCATCATCACTCTTCTTAATCCAGAGAGAGATTGGAATCTCTTATCCCAAC	1321
QY	1249	CAGGCTGTGGCGCCATCATCGGGAAGAGGGGGCACATCAACACAGCTGGCAGATTTC	1308
Db	1322	CAGGCTGTGGGAGCTATCATCGGGAAGAGGGGGCACATCAACACAGCTGGCAGATTTC	1381
QY	1309	GCCGAGGCTCTATCAAGATTGGCCCTCGGAAGGCCCAAGCTGACGGAAGAGATGCTC	1368
Db	1382	GCTGTGCTCTCATCAAGATTGCTCCACAAAGGTCCAGATGTCAGTGAAGAGATGCTC	1441
QY	1369	ATCATCACCGGGCCACCCGGAAGCCCAAGTTCAAAGGCCCAAGGACGATTTTGGAAAATG	1428
Db	1442	ATCATCACCGGGCCACCCGGAAGCCCAAGTTCAAAGGCCCAAGGACGATTTTGGAAAATG	1501
QY	1429	AAAGGGAAGAACTTCTTTTAACCCCAAGAGAAGTGAAGCTGGAAGCGCATATCAAGTG	1488
Db	1502	AAAGGGAAGAACTTCTTTTAACCCCAAGAGAAGTGAAGCTGGAAGCGCATATCAAGTG	1561
QY	1489	CCCTTTTCCACAGCTGCGCGGGGTGATTGGCAAGTGGCAGAACCGTGAACGAATCTGAG	1548
Db	1562	CCCTTTTCCACAGCTGCGCGGGGTGATTGGCAAGTGGCAGAACCGTGAACGAATCTGAG	1621
QY	1549	AACCTTAACAGTGCAGAAAGTCATGTCGCTTGATCCCAAAAGCCAGATGGAATGAGAGAA	1608
Db	1622	AACCTTAACAGTGCAGAAAGTCATGTCGCTTGATCCCAAAAGCCAGATGGAATGAGAGAA	1681
QY	1609	GTCATTCGCAAAATTAATCGGGCACTTTCTTGAACGCAAGTGCACACCGCAAGATCAGG	1668
Db	1682	GTCATTCGCAAAATTAATCGGGCACTTTCTTGAACGCAAGTGCACACCGCAAGATCAGG	1741
QY	1669	GAAATTTGTAACAAGGTGAAGCAGACAGAGCAGAAATACCTCAGGGAGTGCCTCCACAG	1728
Db	1742	GAAATTTGTAACAAGGTGAAGCAGACAGAGCAGAAATACCTCAGGGAGTGCCTCCACAG	1801

QY	1725	CGCAGCAAGTAGAGCTCCCA	CGCAGCAACAGATAGTAGCCCTTCCAA	1788
Db	1802	CGCAGCAAGTAGAGCTCCCA	CGCAGCAACAGATAGTAGCCCTTCCAA	1859
QY	1789	CACCTTGAACGATAGACCCAA	CGCAGCAGCCAGATGGGAGCAACCAAGACAT-C	1847
Db	1860	CACCTTGAACG-ATGAGACCAAA-C	-AGCCAGCAGATGGGAAGCAACCAAGAGCATCC	1915
QY	1848	TGAGCAATAGAAAGTCGTGCGAGGCGGCGAGGAGACTCGCGAGGCGCTGGAAACCCAG	1907	
Db	1916	CGAGAGATGCGAGATCTGAGAGCAGCAGGCGCTGC-----AGACCTCTACATCTGTG	1970	
QY	1908	GAGCCGAGAGAGGGGCGGGAGAGGTCAGCCAGGTTTGGCCAGAACACACG--AGCCCCGCT	1965	
Db	1971	GGATCCAGAGAGGGGCAAGGGAAG--GGCAGGTGTCCAGAAACACGCTTTGGCTGGCC	2027	
QY	1966	CCCGCCCCCAGGGCTTCTGCAGGCTT--CAGCCATCCACTTACCCATCCACTCGGATCT	2023	
Db	2028	CCAGCTTCCCGGCTTCTGCAAGCATACGCCATCCACGCCCATCCACTCAGAT-T	2086	
QY	2024	CTCTGAACTCCCAAGCGCTATCCCTTTAGTTGAATACATAGGGAACGGTTCA	2083	
Db	2087	CTCTCAAGTCCCAAGAGCGCTATCCCTTGGTTGAATACATAGGGAACAGCTCA	2146	
QY	2084	AGCCAGCAAAATGCACACCTTTTTC---TGTGCAAAATCGTCTGTACATGTGT	2139	
Db	2147	AGCAAGCAAAATCTCTAGCCGTTCTTGTGTGTGAAAGTGTCTCTGTACATGTATGT	2206	
QY	2140	ACATATTGAAAAGGGA- GATGTTAGATATGTGGCTGTGGTTTACACAGGCTGCTGC	2198	
Db	2207	ACATATCGAAAGGGAGAGATGTTAAGAAATGTGGCTGTGGTTTACACAGGCTGCTGC	2266	
QY	2199	AGCGTAATATATTTTAGAAATATATATCAATATCACTAATCCATTTTATATC	2258	
Db	2267	AGCGTAATATATTTTAGAAATATATATCAATATCACTAATCCATTTTATATC	2326	
QY	2259	AATTAATATTTTTTTTCTTTTAAAGAGAAACAGGCTTTCTAGACTTTAAAGATA	2318	
Db	2327	AGTATTAATTTTTTTTCTTTTGAAGAGAAACAGGCTTTCTAGACTTTCAAGATA	2386	
QY	2319	AAGTCTTTGG- GAGGCTCACGGTGTAGAGAGAGCTTTGAGGC-CAACCGCAAAAT	2376	
Db	2387	AAGTATGGGTGAGGTCACACTGTATAGAAAGAACTTTGAGGCCACCAACAAAAAT	2446	
QY	2377	CACCCAGAGGGAATTCGTGCGGAAGACACTCAGGCACTTCTGATCACTGTGATG	2436	
Db	2447	CACCTTGAAGGGAATCTGTGTC- AAGGACACTTAATGCGAATGTGGGTCACTGTGCTG	2505	
QY	2437	TCAACAGAGGATACGCTCTCTTTGAAGAGAAACTCTCTACTCTCTCAATGCTGTCA	2496	
Db	2506	TCAACAGAGGGAATACG--TCCTGAGAGAGGGGCCCTAGGCCCTGTCACTGTCTC	2563	
QY	2497	GCTATACACCCATTTCTTT-----TGCTTCAAGGTTTTAA	2534	
Db	2564	CCACACACCCATTCCTCAGTTTCAACAGTTTTTAACTGATGATTTTTGGTTGTTTG	2623	
QY	2535	ACTGTTTTTTGACATACGCTATATAATTCGTCTCTCTCTCTGTTTATCTCTCCCTCC	2594	
Db	2624	TTTTGTTTTTTGACATCTGCTATATATCTGTCTCTC-----CTTATTCCTCTCT	2677	
QY	2595	CTCCCTCCCTTCTCTCATCTCCATCTTTTGAATTTCTCATGCCCATTCAT	2654	
Db	2678	AACCTTCTGTGTCTTTTACCCAAATCTTTTGAATCCCTGTGTCTTTTCTGCTC	2737	
QY	2655	CCCGTATCTAGCAC--CCCCCCCCCAGGCAAGAGGCTCTAGTATCAATCA	2711	
Db	2738	CCCGTATCTATGACACCCCTCCCTCCCCCAAGCAAGAGTCTCTGAATATCATACCA	2797	
QY	2712	CAGAAAAGGAAACAAACGAAACACACAAACCAAGCCTCACTTACACTTGTATCTCAAA	2771	
Db	2798	CACACAGGAAACAAATGCGAACACACACAGCCTCAGCTTATCACTGGTTATCTCAAG	2857	
QY	2772	AGAACAAAGTCAATGTACTGTCTCTAGGTTTTGGAAGAGGAAACAGAAACCAACA	2831	

Db 2858 CGAACAAGAGTGGCGGAGTACTTGTCTTACG-ATTTCGAAAGGAGAAAGCAGATCCCGACA 2916
QY 2832 AACCAACCAATCAACCAACCAACCAACCAACCAATTCACATGAAAGATGTTTGTCTT 2891
Db 2917 AATTAACAGCAGCAGCAACCAACCAAG-AAAATATCCCAAGAAAGAAATGGA----CTCTCT 2971
QY 2892 TTGCATTTTGGTGTATTAAGCCATCAATTTTCAACCAAAATGATTCCTTCTTTAAAAAAA 2951
Db 2972 CTCATCTTGTGTATTAACAGCCATCAACATTTAGTAAATTTCTTCTTTAAAAATTT 3031
QY 2952 AAATGTGAGGAAAGTAGAAATTTTACCAAGGTGTGTGCGCAGGCGCTTAATTCACAGA 3011
Db 3032 AAAGTGCAGGAAATAGCAATTTATACAGGTTGTGTGCGCAGGCGCTTAATTTACAGA 3091
QY 3012 TTTTATTAACGAGAAACACACAGA--AGAGTACTCTAGTG-TTTTTACTCTGAC 3067
Db 3092 TTTTATTAATGAGAAACACACAAATTAAGCTACGCTGAGTGTATTTTACTCTGAC 3151
QY 3068 ACCTTGTCTGTGTCTTTCCTTACAGATTTTGTAAAGCTGATGAGATTTTGTAA 3127
Db 3152 ACCTTGTCTGTGTCTTTCCTTACAGATTTTGTAAAGCTGATGAGATTTTGTAA 3210
QY 3128 TTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3187
Db 3211 TTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3268
QY 3188 TGACAGTCCAAAGTGTGCAACAGCTGTCTGAAATTTGCTTCCGCTACGCAAGACCAAT 3247
Db 3269 CGGAGCGCCAGAGTGTGCGCAGCTCTTCTC---TGTCTTCCGCTACGCAAGAGCTAGCT 3323
QY 3248 GGCCTTCTTTTGGACAAACCTTGAAATGTTATTT 3283
Db 3324 GGCCTTCTTTGAGCAACCTTAATTAATTTT 3359

RESULT 15
AF117107 2010 bp mRNA linear PRI 26-JAN-1999
LOCUS Homo sapiens IGF-II mRNA-binding protein 2 (IMP-2) mRNA, complete
DEFINITION cds.
AF117107
VERSION AF117107.1 GI:4191609
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2010)
AUTHORS Nielsen, J., Christiansen, J., Lykke-Andersen, J., Johnsen, A.H.,
Wewer, U.M. and Nielsen, F.C.
TITLE A family of insulin-like growth factor II mRNA-binding proteins
represses translation in late development
JOURNAL Mol. Cell. Biol. 19 (2), 1262-1270 (1999)
MEDLINE 99108099
PubMed 9891060
REFERENCE 2 (bases 1 to 2010)
AUTHORS Nielsen, J., Christiansen, J., Lykke-Andersen, J., Johnsen, A.H.,
Wewer, U.M. and Nielsen, F.C.
TITLE Direct Substrates
Submitted (30-DEC-1998) Institute of Molecular Biology, University
of Copenhagen, Soelystgade 83H, Copenhagen DK-1307, Denmark
FEATURES
source
1..2010
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..2010
/gene="IMP-2"
47
/gene="IMP-2"
misc_feature
/note="alternative start-codon gives an additional
N-terminal methionine"

CDS
50..1846
/gene="IMP-2"
/note="binds multiple sites in IGF-II mRNA 5'UTR"
/codon_start=1
/product="IGF-II mRNA-binding protein 2"
/protein_id="P40982.1"
/db_xref="GI:4191610"
/translation="MNLVYGNLSPAVATADLRQLFEDRKPLIAGQVLLKSGYAFVY
PDQWALRALETSLGKVELHGIKMEVDYSVKILRSRKIQIRNIPHLQWVLDGLA
QYGTVEVBNVNDPTETAVVNVVATREBAKIAMKLSGHQFENYSPKISYIPBEVS
SPSPQQAORSDHRSREGARPGSTSGAROIDPLRLILVPOFGALITIGKGLATKNI
TKQFQSRVDIHRKNSGAKRPVTIHAFTBETSPACSMILIEIMKEDETKLAERIP
KILAHNLVGRLLIGKEGNLKEKTHETGKTTISLDPLSTYNDERITTYGTYVACA
SAEIEIMKLEAFENEDMLAVNOQANLIPGLNSALGIFSTGLSVSPFAGKAPPA
APYHFTTHSGYFSLYPHQPGFPFPHHSYRDEIVNLPTQAVGAIPKKGAKIK
OLARFAGSIRIKIADPAGPVSERWVITGPPEAPGAKOGRIFGKLEENFENPKBEVK
LEAATRPSSSTAGSIVIGKGTVELNLNLSAEYIVPRDQPDDEMEVITVILGHFFA
SQTAKRKIREIVQVKKQEQKRPQGVASQRK"

Query Match 52.1%; Score 1711.8; DB 9; Length 2010;
Best Local Similarity 92.9%; Pred. No. 0;
Matches 1858; Conservative 0; Mismatches 12; Indels 129; Gaps 1;

QY 35 GCCGGGGAGCCGCGGCTCTTCGGGGAAGAAGCATATGAAACAGCTTTATCATCGGGA 94
Db 12 GCCGTAGGAGACCGCGGCTCAGCGGGAAGAAGCATATGAAACAGCTTTATCATCGGGA 71
QY 95 ACCTGAGCCCGCGCTCACCAGCAAGCTCCGCAAGCTCTTGGGAGACAGAAAGCTGC 154
Db 72 ACCTGAGCCCGCGCTCACCAGCAAGCTCCTCCGCAAGCTCTTGGGAGACAGAAAGCTGC 131
QY 155 CCTTGGCGGAGCAAGTCTCTGTAAGTCCGCTACGCTTGTGTGACTACCCGACAGA 214
Db 132 CCTTGGCGGAGCAAGTCTCTGTAAGTCCGCTACGCTTGTGTGACTACCCGACAGA 191
QY 215 ACTGGGCAATCCGCGCATCGAGACCTCTCGGGTAAGTGAATTTGATGGGAAATGA 274
Db 192 ACTGGGCAATCCGCGCATCGAGACCTCTCGGGTAAGTGAATTTGATGGGAAATGA 251
QY 275 TGGAGTGTATTAACACAGCTCTTAATAAAGCTAAGAGCAGAAATTCAGATTCGAAACA 334
Db 252 TGGAGTGTATTAACACAGCTCTTAATAAAGCTAAGAGCAGAAATTCAGATTCGAAACA 311
QY 335 TCCCTCTCACCCTGAGTGGAGGTGTGATGGAATTTGCTCAATATGGAACAGTGG 394
Db 312 TCCCTCTCACCCTGAGTGGAGGTGTGATGGAATTTGCTCAATATGGAACAGTGG 371
QY 395 AGAATGTGACACAGTCAACACAGACACGAAACCGCGCTTGTCAACGTCACATATGGA 454
Db 372 AGAATGTGACACAGTCAACACAGACACGAAACCGCGCTTGTCAACGTCACATATGGA 431
QY 455 CAAGGAAGAAGCAAAATATAGCATGAGAGACTAAGCGGCACTGATTGAGAACTACT 514
Db 432 CAAGGAAGAAGCAAAATATAGCATGAGAGACTAAGCGGCACTGATTGAGAACTACT 491
QY 515 CCTTCAAGATTTCTTCAATCCCGATGAAAGAGTGAAGTCTCCCTTGCCTCTCAGCGAG 574
Db 492 CTTTCAAGATTTCTTCAATCCCGATGAAAGAGTGAAGTCTCCCTTGCCTCTCAGCGAG 551
QY 575 CCGAGCGTGGGACCACTCTTCCCGGAGCAAGCGCCCTTGGGGGCACTTCCAGG 634
Db 552 CCGAGCGTGGGACCACTCTTCCCGGAGCAAGCGCCCTTGGGGGCACTTCCAGG 611
QY 635 CCAGACAGATGATTTCCGCTCGGATCTGTGCCACCCAGTTTGTGTGATCA 694
Db 612 CCAGACAGATGATTTCCGCTCGGATCTGTGCCACCCAGTTTGTGTGATCA 671
QY 695 TCGGAAAGAGGAGCTTGAACATTAAGAAACATCACTAAGCAGACCCAGTCCCGGTAATA 754
Db 672 TCGGAAAGAGGAGCTTGAACATTAAGAAACATCACTAAGCAGACCCAGTCCCGGTAATA 731
QY 755 TCATAGAAAAGAGAACTTGAGCTGAGAGAGAGAGCTGTACATCATCATGCCACCCAG 814

Db 732 TCCATAGAAAAGAGAACTCTGGAGCTGCAGAGAACTCTGTCACTTCATGCAACCCAG 791
 QY 815 AGGGGACTTTTGAAGCATGCGCATGATTTTGAATATATGCAAGAAAGAGCATTTGA 874
 Db 792 AGGGGACTTTTGAAGCATGCGCATGATTTTGAATATATGCAAGAAAGAGCATTTGA 851
 QY 875 CCAAACTAGCCGGAAGAGATTCCTCTGAAAATCTTTGGCACAACATGCGCTTGGTGGAAAG 934
 Db 852 CCAAACTAGCCGGAAGAGATTCCTCTGAAAATCTTTGGCACAACATGCGCTTGGTGGAAAG 911
 QY 935 TGATTTGAAAAAGAGAGAGAAATTTGAAGAAATTTGAACATGAAACAGGAGCAAGATTA 994
 Db 912 TGATTTGAAAAAGAGAGAGAAATTTGAAGAAATTTGAACATGAAACAGGAGCAAGATTA 971
 QY 995 CAATCTCATCTTTTGCAGATTTTGCAGATTAACACCCGGAAGAAACCATCTCTGTAAG 1054
 Db 972 CAATCTCATCTTTTGCAGATTTTGCAGATTAACACCCGGAAGAAACCATCTCTGTAAG 1031
 QY 1055 GCACAGTTGAGGCTGTGCGCAGTGTGATAGATTAATGAGAGAGCTGGTGAAGGCT 1114
 Db 1032 GCACAGTTGAGGCTGTGCGCAGTGTGATAGATTAATGAGAGAGCTGGTGAAGGCT 1091
 QY 1115 TTGAAAATGATATGCTGGCTGTTAAC----- 1140
 Db 1092 TTGAAAATGATATGCTGGCTGTTAACCAACAAGCCAAATGTGATCCAGGGTTGAACCTCA 1151
 QY 1141 ----- 1140
 Db 1152 GCGCATTTGGCATCTTTTCAACAGAGACTGTCCGTATCTCAACAGAGGGCCCCGCG 1211
 QY 1141 -----ACCCACTCGGAGTACTCTTCCAGCC 1165
 Db 1212 GAGTCCCCCGCGCTGCCCTTACCAACCCCTTCACTACCCACTCCGAAATCTTCTCAAGCC 1271
 QY 1166 TGTACCCCATCAACAGATTTGCGCCGTTCGCGATCACTCTTATCCAGAGCAGAGA 1225
 Db 1272 TGTACCCCATCAACAGATTTGCGCCGTTCGCGATCACTCTTATCCAGAGCAGAGA 1331
 QY 1226 TTGTGAATCTCTTCAATCCCAACCGAGCTGTGGGCGCATCATGCGGAGAGAGGGGCGAC 1285
 Db 1332 TTGTGAATCTCTTCAATCCCAACCGAGCTGTGGGCGCATCATGCGGAGAGAGGGGCGAC 1391
 QY 1286 ACATCAAAACAGCTGGCAGATTTGCGCGAGCTCTATCAAGATTTGCCCTCGGAGAGGCC 1345
 Db 1392 ACATCAAAACAGCTGGCAGATTTGCGCGAGCTCTATCAAGATTTGCCCTCGGAGAGGCC 1451
 QY 1346 CAGACGTCAGCGAAGAGATGATCATCAACCGGCGCACCGGAAAGCCAGTTCAAGGCC 1405
 Db 1452 CAGACGTCAGCGAAGAGATGATCATCAACCGGCGCACCGGAAAGCCAGTTCAAGGCC 1511
 QY 1406 AGGGAAGAGATCTTTGGGAAAATGAAAGAGAAAATCTTTTAAACCCCAAGAAAGATGA 1465
 Db 1512 AGGGAAGAGATCTTTGGGAAAATGAAAGAGAAAATCTTTTAAACCCCAAGAAAGATGA 1571
 QY 1466 AGCTGGAAGGCAATATCAGAGTGCCTTCCACAGCTGGCGGAGATTTGCGAAAGTG 1525
 Db 1572 AGCTGGAAGGCAATATCAGAGTGCCTTCCACAGCTGGCGGAGATTTGCGAAAGTG 1631
 QY 1526 GCAAGACCGTGAAGCAATGCAAGACTTAAACAGTGCAGAAATCATGTCCTCGTACCC 1585
 Db 1632 GCAAGACCGTGAAGCAATGCAAGACTTAAACAGTGCAGAAATCATGTCCTCGTACCC 1691
 QY 1586 AAATCGCCAGATGAATAATGAGGAAGTGAATGTCAGAAATTAATCGGCACTTCTTTGCTAGCC 1645
 Db 1692 AAATCGCCAGATGAATAATGAGGAAGTGAATGTCAGAAATTAATCGGCACTTCTTTGCTAGCC 1751
 QY 1646 AGACTGCAAGCGCAGAGATCAGGGAATAATGTACAACAGGTGAAGCAGAGCAGAGAAAT 1705
 Db 1752 AGACTGCAAGCGCAGAGATCAGGGAATAATGTACAACAGGTGAAGCAGAGCAGAGAAAT 1811
 QY 1706 ACCCTGAGGAGTGCCTCAACAGCGCAAGAGTGAAGCTTCCACAGGCAACAGCAAAACA 1765

Db 1812 ACCCTGAGGAGTGCCTCAACAGCGCAGAGCAAGTGAAGCTCCCAAGGCAACCAAAACA 1871
 QY 1766 ACGGATGATATGAGCCCTTCCAAACCTGACAGAAATGAGACCAACGCGAGCCAGCAT 1825
 Db 1872 ACGGATGATATGAGCCCTTCCAAACCTGACAGAAATGAGACCAACGCGAGCCAGCAT 1931
 QY 1826 CGGAGGCAAAACCAAGACCATCTGAGGAATGAGAAATCTGTGGAGGCGGCGCAAGGACTCT 1885
 Db 1932 CGGAGGCAAAACCAAGACCATCTGAGGAATGAGAAATCTGTGGAGGCGGCGCAAGGACTCT 1991
 QY 1886 GCCGAGGCCCTGAGAACCC 1904
 Db 1992 GCCGAGGCCCTGAGAACCC 2010

Search completed: July 24, 2004, 13:04:44
 Job time : 12667.7 secs

This Page Blank (uspto)

The present sequence represents an alternative form of a cancer associated antigen gene designated KOC-3. The specification also describes a cancer associated antigen designated CT7. The CT7 polynucleotide was isolated from SK-MEL-37 melanoma cells. The polypeptide has some homology with MAGE-10, limited to about 210 carboxy terminal amino acids. The amino terminal of the protein has a repetitive pattern, with repeats rich in serine, proline, glutamine and leucine, and an almost invariable core of the peptide given in AA443877. The CT7 polypeptide can be processed to peptides which provoke lysis by cytolytic T cells. The polynucleotides and polypeptides can be used for treating a cancerous condition and screening for or diagnosing cancerous conditions. The cancer associated antigens can be used as an immunogenic or vaccine composition with an adjuvant, e.g. a cytokine, a saponin, or granulocyte

CC macrophage-colony stimulating factor (GM-CSF)
XX Sequence 3283 BP, 945 A, 833 C, 779 G, 725 T, 0 U, 1 Other;
SQ Best Local Similarity 100.0%; Score 3282; DB 3; Length 3283;
Query Match 100.0%; Pred. No. 0;
Matches 3283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCAGCGAGAGAGCGAGAGAGCGCGGATCCGGCCCGGGAGCGCGGGCTCCGGGG 60
DB 1 GGCAGCGAGAGAGAGAGAGCGCGGATCCGGCCCGGGAGCGCGGGCTCCGGGG 60
QY 61 AAGNACGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
DB 61 AAGNACGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
QY 121 GACCTCCGGAGCTCTTTGGGGACAGAAAGCTGCCCTGGCGGACAGGCTCTGTGAG 180
DB 121 GACCTCCGGAGCTCTTTGGGGACAGAAAGCTGCCCTGGCGGACAGGCTCTGTGAG 180
QY 181 TCCGGCTACGCTTGTGACTACCCCGACAGAACTGGGCGATCCGGCCATCGAGACC 240
DB 181 TCCGGCTACGCTTGTGACTACCCCGACAGAACTGGGCGATCCGGCCATCGAGACC 240
QY 241 CTCTGGGTAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 241 CTCTGGGTAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 301 AAGCTAAGGAGCAGGAAATTCAGATTCGAAATCCTCTCTCACTGCACTGGAGGTG 360
DB 301 AAGCTAAGGAGCAGGAAATTCAGATTCGAAATCCTCTCTCACTGCACTGGAGGTG 360
QY 361 TTGGATGACTTTGGCTCAATATGAGGACAGTGGAGAAATGAGAAATGCAACAGAC 420
DB 361 TTGGATGACTTTGGCTCAATATGAGGACAGTGGAGAAATGAGAAATGCAACAGAC 420
QY 421 TGGATGACTTTGGCTCAATATGAGGACAGTGGAGAAATGAGAAATGCAACAGAC 480
DB 421 TGGATGACTTTGGCTCAATATGAGGACAGTGGAGAAATGAGAAATGCAACAGAC 480
QY 481 GAGAACTAAGCGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
DB 481 GAGAACTAAGCGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
QY 541 GAAAGGTGAGTCTCCCTTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCG 600
DB 541 GAAAGGTGAGTCTCCCTTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCG 600
QY 601 GAGCAAGGCGCGCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCT 660
DB 601 GAGCAAGGCGCGCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCT 660
QY 661 ATCTGGTCCCAACCGATTTGTTGGTCCATATCGGAAAGAGGCTTGAACATAAG 720
DB 661 ATCTGGTCCCAACCGATTTGTTGGTCCATATCGGAAAGAGGCTTGAACATAAG 720
QY 721 AATCATCTAAGACAGCCAGTCCCGGATGATGATGATGATGATGATGATGATGATGAT 780
DB 721 AATCATCTAAGACAGCCAGTCCCGGATGATGATGATGATGATGATGATGATGATGAT 780
QY 781 GCGAGAGGCTGTCACTCATGCGACCCAGAGGAGCTTGTGAAGCATGCCGATG 840
DB 781 GCGAGAGGCTGTCACTCATGCGACCCAGAGGAGCTTGTGAAGCATGCCGATG 840
QY 841 ATTCTTGAATCTGACAGAAAGAGGAGATGAGCCTAAGCTGAGCGAAGATTCCTCTG 900
DB 841 ATTCTTGAATCTGACAGAAAGAGGAGATGAGCCTAAGCTGAGCGAAGATTCCTCTG 900
QY 901 AAAATCTTGGACACATGCTTGTGTGAGAACTGATGATGATGATGATGATGATGATGAT 960
DB 901 AAAATCTTGGACACATGCTTGTGTGAGAACTGATGATGATGATGATGATGATGATGAT 960
QY 961 AAGAAATGAACTGAACAGGAGCAGATTAACATCAATCTTGTGAGATTTGAGC 1020

DB 961 AAGAAATGAACTGAACAGGAGCAGATTAACATCAATCTTGTGAGATTTGAGC 1020
QY 1021 AATATCAACCCGGAAGAGACCATCATCTGTGAAGGACAGATGTAGGCTGTGCCAGTCT 1080
DB 1021 AATATCAACCCGGAAGAGACCATCATCTGTGAAGGACAGATGTAGGCTGTGCCAGTCT 1080
QY 1081 GAGATGAGATTAAGAAAGCTGTGAGGCTTTGAAATGATATGCTGTGCTTAAAC 1140
DB 1081 GAGATGAGATTAAGAAAGCTGTGAGGCTTTGAAATGATATGCTGTGCTTAAAC 1140
QY 1141 ACCCATCTCCGATCTTCTCCAGCTTGAACCCCATCAACGTTTGGCCCGTCCCGAT 1200
DB 1141 ACCCATCTCCGATCTTCTCCAGCTTGAACCCCATCAACGTTTGGCCCGTCCCGAT 1200
QY 1201 CATCACTCTTATCCAGACAGAGATGATGATGATGATGATGATGATGATGATGATGAT 1260
DB 1201 CATCACTCTTATCCAGACAGAGATGATGATGATGATGATGATGATGATGATGATGAT 1260
QY 1261 GCCATCATCGGAGAGAGGCGCACATCAACAGCTGCGAGATTCGCGGAGCTCT 1320
DB 1261 GCCATCATCGGAGAGAGGCGCACATCAACAGCTGCGAGATTCGCGGAGCTCT 1320
QY 1321 ATCAAGATTTGCCCTTCCGGAAGGCCCAAGCTCAGCGAAAGGATGATCATCAACCGG 1380
DB 1321 ATCAAGATTTGCCCTTCCGGAAGGCCCAAGCTCAGCGAAAGGATGATCATCAACCGG 1380
QY 1381 CCACCGAAGCCCACTTCAAGGCCAGGACGATCTTTGGAAACTGAAAGAGGAAAC 1440
DB 1381 CCACCGAAGCCCACTTCAAGGCCAGGACGATCTTTGGAAACTGAAAGAGGAAAC 1440
QY 1441 TTCTTTAAACCCCAAGAAAGATGAAAGCTGAAAGGCAATATCAAGTGCCTCTTCCA 1500
DB 1441 TTCTTTAAACCCCAAGAAAGATGAAAGCTGAAAGGCAATATCAAGTGCCTCTTCCA 1500
QY 1501 GCTGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
DB 1501 GCTGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
QY 1561 GCGAAGATCATGATGCTGCTGTAACCAACGCGAGATGAAGATGATGATGATGATGAT 1620
DB 1561 GCGAAGATCATGATGCTGCTGTAACCAACGCGAGATGAAGATGATGATGATGATGAT 1620
QY 1621 ATTATCGGCACTTCTTCTTACAGCACTGCAACGCGAAGATTCAGGAAATTTGACAA 1680
DB 1621 ATTATCGGCACTTCTTCTTACAGCACTGCAACGCGAAGATTCAGGAAATTTGACAA 1680
QY 1681 CAGGTGAACAGACAGAGAAATACCTTCAAGGAGTGCCTTCAACGCGACCAAGTGA 1740
DB 1681 CAGGTGAACAGACAGAGAAATACCTTCAAGGAGTGCCTTCAACGCGACCAAGTGA 1740
QY 1741 GGTCTCCACAGGACAGCAAAACCAACGATGATGATGATGATGATGATGATGATGATGAT 1800
DB 1741 GGTCTCCACAGGACAGCAAAACCAACGATGATGATGATGATGATGATGATGATGATGAT 1800
QY 1801 TGAACCTAAGCAGACGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860
DB 1801 TGAACCTAAGCAGACGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860
QY 1861 GTCTGCGAGGCGGACAGGAGCTCTGCGAGGCTCTGAGAACCCCAAGGCGCGAGAGG 1920
DB 1861 GTCTGCGAGGCGGACAGGAGCTCTGCGAGGCTCTGAGAACCCCAAGGCGCGAGAGG 1920
QY 1921 GCGGAGAGGTCAAGCAGATTTGCCAAGAACACCGAGCCCTCCGCTCCGCTCCGCTCC 1980
DB 1921 GCGGAGAGGTCAAGCAGATTTGCCAAGAACACCGAGCCCTCCGCTCCGCTCCGCTCC 1980
QY 1981 TTCTGACGCTTCAAGCATCACTTCAACCTCACTGATCTCTCTGAACTCCAGCA 2040
DB 1981 TTCTGACGCTTCAAGCATCACTTCAACCTCACTGATCTCTCTGAACTCCAGCA 2040
QY 2041 CGCTATCCCTTTAGTGAATTAATAGGAGACGTTTCAAGCCCAAGAAATGAGC 2100


```
Db      2041 CGGTATCCCTTTAGTGAACATGAGTGAACGCTGTTCAAGCCAAAGCAAAATGCAC 2100
QY      2101 ACCCTTTTCTGTGGCAAAATCGTCTGTACATGTTGATACATATTGAAGAGGAAG 2160
Db      2101 ACCCTTTTCTGTGGCAAAATCGTCTGTACATGTTGATACATATTGAAGAGGAAG 2160
QY      2161 TTAAGATATGTGGCTGTGGTTACAAGAGTGCCGACAGCGTAATATATTTAGAAAT 2220
Db      2161 TTAAGATATGTGGCTGTGGTTACAAGAGTGCCGACAGCGTAATATATTTAGAAAT 2220
QY      2221 AATATATCAATTAATCTCAACTAATCTCAATTTTAAATCAATTAATTTTCTTT 2280
Db      2221 AATATATCAATTAATCTCAACTAATCTCAATTTTAAATCAATTAATTTTCTTT 2280
QY      2281 TTAAGAGAAAGCAGGCTTTTGTAGACTTTAAAGAAATTAAGCTTTGGAGGTCTCAGG 2340
Db      2281 TTAAGAGAAAGCAGGCTTTTGTAGACTTTAAAGAAATTAAGCTTTGGAGGTCTCAGG 2340
QY      2341 TGTAGAGAGAGCTTTGAGGCAACCCGACAAATTCACCCAGAGGAAATCTCGTGGGA 2400
Db      2341 TGTAGAGAGAGCTTTGAGGCAACCCGACAAATTCACCCAGAGGAAATCTCGTGGGA 2400
QY      2401 AGGACACTCAGGCGAGTCTGTGATACCTGTGTATGTCAAGAGAGATACCGTCTCT 2460
Db      2401 AGGACACTCAGGCGAGTCTGTGATACCTGTGTATGTCAAGAGAGATACCGTCTCT 2460
QY      2461 TGAAGAGAAACTCTGTCACTCCTCATGCTGTAGCTCAATACCCATTTCTTTGC 2520
Db      2461 TGAAGAGAAACTCTGTCACTCCTCATGCTGTAGCTCAATACCCATTTCTTTGC 2520
QY      2521 TTCAAGGTTTAAACTGTTTTTGTGATCTGATATTAATTTCTGTCTCTCTGT 2580
Db      2521 TTCAAGGTTTAAACTGTTTTTGTGATCTGATATTAATTTCTGTCTCTCTGT 2580
QY      2581 TATCTTCCTCTCCTCTCCTCTCCTCTCTCTCTCTCTCAATCTCAATCTTTGAAATTCCTCAT 2640
Db      2581 TATCTTCCTCTCCTCTCCTCTCCTCTCTCTCTCTCTCAATCTTTGAAATTCCTCAT 2640
QY      2641 CCCCTCATCTCAATCCCGTATCTAGGACACCCCGCCCGCCAGGCAAGAGCTCTGA 2700
Db      2641 CCCCTCATCTCAATCCCGTATCTAGGACACCCCGCCCGCCAGGCAAGAGCTCTGA 2700
QY      2701 GTATCATATCAACAAAGAGAAAGCAAGCAACAAACAGCTCTCACTTACCTT 2760
Db      2701 GTATCATATCAACAAAGAGAAAGCAAGCAACAAACAGCTCTCACTTACCTT 2760
QY      2761 GGTACTCTCAAAAGAAAGAGTCAATGTAATCTGTCTTCTAGCGTTTGGAAAGAGAAACA 2820
Db      2761 GGTACTCTCAAAAGAAAGAGTCAATGTAATCTGTCTTCTAGCGTTTGGAAAGAGAAACA 2820
QY      2821 GGAACCCACCAACCAACCAATCAACCAACCAAGAAATTTCCAAATGAAAGATGT 2880
Db      2821 GGAACCCACCAACCAACCAATCAACCAACCAAGAAATTTCCAAATGAAAGATGT 2880
QY      2881 AATTTGTCTTTTGCATTTTGTGTATAGCCATCAATTTACGAAATGATTTCTCTTC 2940
Db      2881 AATTTGTCTTTTGCATTTTGTGTATAGCCATCAATTTACGAAATGATTTCTCTTC 2940
QY      2941 TTTAAAGAAAGAAATGTGAGAGAAAGTAAATTTACCAAGTTTGGCCAGGCGGT 3000
Db      2941 TTTAAAGAAAGAAATGTGAGAGAAAGTAAATTTACCAAGTTTGGCCAGGCGGT 3000
QY      3001 AAATTCACAGATTTTAAACGAGAAACACACAGAGAGTACTCTCAGGTCTTTTA 3060
Db      3001 AAATTCACAGATTTTAAACGAGAAACACACAGAGAGTACTCTCAGGTCTTTTA 3060
QY      3061 CCTCAGACCTTGTCTGTGTCTTGTAGAGATTTTGAAGTGTATGTTGAGAGAT 3120
Db      3061 CCTCAGACCTTGTCTGTGTCTTGTAGAGATTTTGAAGTGTATGTTGAGAGAT 3120
QY      3121 TTTTATTTTATTTTAAATGAGTTGAGAAATTAAGATCAACTGCGACGCTG 3180
Db      3121 TTTTATTTTATTTTAAATGAGTTGAGAAATTAAGATCAACTGCGACGCTG 3180
```

```
QY      3181 GAGAAAGTGACAGTCCAAAGTGTGCAACAGCTGTTCTGAATTTCTTCGCTAGCCAGAA 3240
Db      3181 GAGAAAGTGACAGTCCAAAGTGTGCAACAGCTGTTCTGAATTTCTTCGCTAGCCAGAA 3240
QY      3241 CCNATATGCGCTTTTGGACAAACCTTGAAGAAATTTATTT 3283
Db      3241 CCNATATGCGCTTTTGGACAAACCTTGAAGAAATTTATTT 3283

RESULT 2
ID      AAS70982 standard; cDNA, 3667 BP.
XX
XX      AAS70982;
AC
XX
XX      13-FEB-2002 (first entry)
DT
XX
XX      DNA encoding novel human diagnostic protein #6786.
DE
XX
XX      Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW      food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS
XX      Homo sapiens.
EN      WO200175067-A2.
XX
XX      11-OCT-2001.
PD
XX
XX      30-MAR-2001; 2001WO-US008631.
PF
XX
XX      31-MAR-2000; 2000US-00540217.
PR      23-AUG-2000; 2000US-00649167.
XX
XX      (HXS-) HXSEQ INC.
PA
XX
XX      Dmanac RT, Liu C, Tang YT,
PI
XX
XX      WPI; 2001-639362/73.
DR      P-PSDB; ABG06795.
XX
XX
PT      New isolated polynucleotide and encoded polypeptides, useful in
PT      diagnostics, forensics, gene mapping, identification of mutations
PT      responsible for genetic disorders or other traits and to assess
PT      biodiversity.
XX
XX      Claim 1; SEQ ID NO 6786; 103pp; English.
XX
XX      The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC      sequences. (I) is useful as hybridisation probes, polymerase chain
CC      reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC      and in recombinant production of (II). The polynucleotides are also used
CC      in diagnostics as expressed sequence tags for identifying expressed
CC      genes. (I) is useful in gene therapy techniques to restore normal
CC      activity of (II) or to treat disease states involving (II). (II) is
CC      useful for generating antibodies against it, detecting or quantitating a
CC      polypeptide in tissue, as molecular weight markers and as a food
CC      supplement. (II) and its binding partners are useful in medical imaging
CC      of sites expressing (II). (I) and (II) are useful for treating disorders
CC      involving aberrant protein expression or biological activity. The
CC      polypeptide and polynucleotide sequences have applications in
CC      diagnostics, forensics, gene mapping, identification of mutations
CC      responsible for genetic disorders or other traits to assess biodiversity
CC      and to produce other types of data and products dependent on DNA and
CC      amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC      coding sequences of the invention. Note: The sequence data for this
CC      patent did not appear in the printed specification, but was obtained in
CC      electronic format directly from WIPO at
CC      ftp.wipo.int/pub/published_pct_sequences
XX
XX      Sequence 3667 BP; 1081 A; 882 C; 846 G; 858 T; 0 U; 0 Other;
SQ
Query Match      99.1%; Score 3253.8; DB 5; Length 3667;
```

	Best Local Similarity	99.8%	Pred: No. 0:	Mismatches	0;	Mismatches	3;	Indels	2;	Gaps	2;
QY	3	CAGCGAGGAGGGCAGAGAGCCGCCGGGTACCGGCGCGGGGAGCCCGCGGGCTTTCGGGGAA	62								
Db	369	CAACGAGAGGAGCGAGGAGCGCGGGGTACCGGGCGCGGGGAAGCCCGGGGCTTCGGGGAAA	428								
QY	63	GAGACGGATATGAAACAAGCTTTAACAATCGGGAAACTGAGCCCCGCCTGACCGCCGAGA	122								
Db	429	GAGCGATATATGAACAAGCTTTAACAATCGGGAACTGAGCCCCGCCTGACCGCCGAGA	488								
QY	123	CCTCCGCGACGCTCTTTGGGGACAAGAGCTGCCCTTGCGGGACAAGTCTCTGTAAGTC	182								
Db	489	CCTCCGCGAGCTCTTTGGGGACAAGAGCTGCCCTTGCGGGACAAGTCTCTGTAAGTC	548								
QY	183	CGGCTTAGCCTTCGTGGACTAACCCCACAGAACTGGGGCATCGGGCATGAGACCCT	242								
Db	549	CGGCTTAGCCTTCGTGGACTAACCCCACAGAACTGGGGCATCGGGCATGAGACCCT	608								
QY	243	CTCGGGTAAATGGAAATTGCAATGGGAAATCAATGAAAGTTGATTACTAGTCTTAAAAA	302								
Db	609	CTCGGGTAAATGGAAATTGCAATGGGAAATCAATGAAAGTTGATTACTAGTCTTAAAAA	668								
QY	303	GCTAAGAGCAGGAAATTCAGATTGAAAATCTCCCTCACCTGCAATGGGAGAGTGT	362								
Db	669	GCTAAGAGCAGGAAATTCAGATTGAAAATCTCCCTCACCTGCAATGGGAGAGTGT	728								
QY	363	GGATGGAATTTTGGCTCAATATGGGACAGTGGAGAAATGTGAAACAAAGTCAACACAGAC	422								
Db	729	GGATGGAATTTTGGCTCAATATGGGACAGTGGAGAAATGTGAAACAAAGTCAACACAGAC	788								
QY	423	AGAAACCCGCGTTGTCAAAGTCAATATGCAACAGAGAGAGAGAAATTAACCTATGGA	482								
Db	789	AGAAACCCGCGTTGTCAAAGTCAATATGCAACAGAGAGAGAGAAATTAACCTATGGA	848								
QY	483	GAACTAAGCGGGGACATCAATTTGAGAAATACTCCCTCAAGATTTCCAAATCCCGGATGA	542								
Db	849	GAACTAAGCGGGGACATCAATTTGAGAAATACTCCCTCAAGATTTCCAAATCCCGGATGA	908								
QY	543	AAGAGTAGACTCCCTTCGCCCCCTCAGCGAGCCAGGTGGGGACCACTTTCGGGGAA	602								
Db	909	AAGAGTAGACTCCCTTCGCCCCCTCAGCGAGCCAGGTGGGGACCACTTTCGGGGAA	968								
QY	603	GGAAGGCCACGCCCCCTGGGGGACATTTCTCAAGGCCAGACAGATTTTCCCGCGGGAT	662								
Db	969	GGAAGGCCACGCCCCCTGGGGGACATTTCTCAAGGCCAGACAGATTTTCCCGCGGGAT	1028								
QY	663	CCGAGCCCCACCAAGTTTGTGTGGGCATCATGSGAAAGAGGGGCTTGACATAAAGAA	722								
Db	1029	CCGAGCCCCACCAAGTTTGTGTGGGCATCATGSGAAAGAGGGGCTTGACATAAAGAA	1088								
QY	723	CATCACTAAGCAGAACCCAAGTCCCGGTATGATATTCATAGAAAAGAGAACTTGSAGTGC	782								
Db	1089	CATCACTAAGCAGAACCCAAGTCCCGGTATGATATTCATAGAAAAGAGAACTTGSAGTGC	1148								
QY	783	AGAGAGGCTGTACCATCATGCAACCCCAAGGGGACTTCTGAAGCATCCGCATGAT	842								
Db	1149	AGAGAGGCTGTACCATCATGCAACCCCAAGGGGACTTCTGAAGCATCCGCATGAT	1208								
QY	843	TCTTGAATCATGCAAGAAAGAGGCAAGATGACAAACTAGCCGAAAGATTTCTCTGAA	902								
Db	1209	TCTTGAATCATGCAAGAAAGAGGCAAGATGACAAACTAGCCGAAAGATTTCTCTGAA	1268								
QY	903	AATCTTGGCACACAAATGGCTTGGTGGAGACATGATGGAAAAGAGGCAAGAAATTGAA	962								
Db	1269	AATCTTGGCACACAAATGGCTTGGTGGAGACATGATGGAAAAGAGGCAAGAAATTGAA	1328								
QY	963	GAAATTTGAACATGAAGACAGGGACCAAGATTAACAATCTCATCTTTGCAAGATTTGAGCAT	1022								
Db	1329	GAAATTTGAACATGAAGACAGGGACCAAGATTAACAATCTCATCTTTGCAAGATTTGAGCAT	1388								
QY	1023	ATTCAAACCCGAAAGAACATCATCTGTGAAGGSCACAGTTGAGGCTGTGCCATGTCTGA	1082								

Db	1389	ATCACACC	GGAAAGAAACCATCACTGTGTAAGGGCA	CAGTTGAGGCGCTGTCCAGTCTGA	1448
Qy	1083	GATTAGAATTATGAAGAACTGTGTGAGCGCTTTGAA	AAATGATATGTCTGGCTTTTAA	CAC	1142
Db	1449	GATTAGAGATTATGAAGAACTGTGTGAGGCGCTTTGAA	AAATGATATGTCTGTGTTAAC	CAC	1508
Qy	1143	CCACTCCGGAAATCTTCCAGGCTGTAAACCCCATCA	CAGTTTGGCCGCTTCCCGCATCA		1202
Db	1509	CCACTCCGGAAATCTTCCAGGCTGTAAACCCCATCA	CAGTTTGGCCGCTTCCCGCATCA		1568
Qy	1203	TCACTCTTATCCAGAGCAGAGAGATTGTGATCTTT	CATCCCAACCCAGGCTGTGGGCGC		1262
Db	1569	TCACTCTTATCCAGAGCAGAGAGATTGTGATCTTT	CATCCCAACCCAGGCTGTGGGCGC		1628
Qy	1263	CATCATCGGGAAAGAAAGGGGGCACATCAATCAAC	AGCTGGCGAGATTTGGCGCGAGGCTCTAT		1322
Db	1629	CATCATCGGGAAAGAAAGGGGGCACACATCAACAG	CTGGCGAGATTTGGCGCGAGGCTCTAT		1688
Qy	1323	CAAGATTGGCCCTCGCGGAAAGGGCCACAGCTCA	GCAGGAAAGATGTGTATCATCACCGGGCC		1382
Db	1689	CAAGATTGGCCCTCGCGGAAAGGGCCACAGCTCA	GCAGGAAAGATGTGTATCATCACCGGGCC		1748
Qy	1383	ACCGGAAGCCCAAGTTCAAGGCCCAAGGACGATCT	TTTGGGAACTGAAAGAGAAACTT		1442
Db	1749	ACCGGAAGCCCAAGTTCAAGGCCCAAGGACGATCT	TTTGGGAACTGAAAGAGAAACTT		1808
Qy	1443	CTTTAAACCCCAAGAAAGAAAGTGAAGCTGGAAG	AGGCAATATCAGATGGCCCTCTTCCACAGC		1502
Db	1809	CTTTAAACCCCAAGAAAGAAAGTGAAGCTGGAAG	AGGCAATATCAGATGGCCCTCTTCCACAGC		1868
Qy	1503	TGGCCGGGGTATTTGGCAAAAGTGTGCAAGCCGT	GAAGCGAACTTGACGAACTTAAACGTTGC		1562
Db	1869	TGGCCGGGGTATTTGGCAAAAGTGTGCAAGCCGT	GAAGCGAACTTGACGAACTTAAACGTTGC		1928
Qy	1563	AGAAATCATGTGTCTGTGACCAACGCCCAATGA	AAATTAAGGAAAGTATGTCTAGAT		1622
Db	1929	AGAAATCATGTGTCTGTGACCAACGCCCAATGA	AAATTAAGGAAAGTATGTCTAGAT		1988
Qy	1623	TATGTGGGACATTTCTTTGCTAGCCGACCTGCA	CAGGCAAGATCAGGAAATTTGTACAA	CA	1682
Db	1989	TATGTGGGACATTTCTTTGCTAGCCGACCTGCA	CAGGCAAGATCAGGAAATTTGTACAA	CA	2048
Qy	1683	GGTGAACAGACAGGAGCAGAATACCTCAGGAGT	CGCTTCACAGCGCAGCAAGTGAAG		1742
Db	2049	GGTGAACAGACAGGAGCAGAATACCTCAGGAGT	CGCTTCACAGCGCAGCAAGTGAAG		2108
Qy	1743	CTCCACAGGCAACGCAAAACAAACGATGAATGA	AGCCCTTCCAAACCTTGAACGAATG		1802
Db	2109	CTCCACAGGCAACGCAAAACAAACGATGAATGA	AGCCCTTCCAAACCTTGAACGAATG		2168
Qy	1803	AGACCAAAACGACAGCCAGCCAGTGGGGGCAAA	CCAAAGACCATCTGAGGAATGAGAAT		1862
Db	2169	AGACCAAAACGACAGCCAGCCAGTGGGGGCAAA	CCAAAGACCATCTGAGGAATGAGAAT		2228
Qy	1863	CTGCGGAGCGGCGCAGGACTGTGCGAGGCGCT	TGAGAACCCCAAGGGGCTGAGAGGGGC		1922
Db	2229	CTGCGGAGCGGCGCAGGAGCTGTGCGAGGCGCT	TGAGAACCCCAAGGGGCTGAGAGGGGC		2288
Qy	1923	GGGGAAGGTCAAGCGTTTGTCAAGAACCCGAG	CGCCGCTCCCGGCCCCCAAGGGCTT		1982
Db	2289	GGGGAAGGTCAAGCGTTTGTCAAGAACCCGAG	CGCCGCTCCCGGCCCCCAAGGGCTT		2348
Qy	1983	CTGCAAGGTTCAAGCCATCCACTTCAACACTCG	ATCGGATCTCTCTGAACTCCACGACG		2042
Db	2349	CTGCAAGGTTCAAGCCATCCACTTCAACACTCG	ATCGGATCTCTCTGAACTCCACGACG		2408
Qy	2043	CTATCCCTTTTATGTAACATAAGGTGAACGTGT	TAAAGCCAAAGCAAAATGCAAC		2102
Db	2409	CTATCCCTTTTATGTAACATAAGGTGAACGTGT	TAAAGCCAAAGCAAAATGCAAC		2468
Qy	2103	CCTTTTTCTGTGGCAAAATCGTCTCTGTACATGT	GTGTACATTTAAGAAAGGAAAGATGTT		2162
Db	2469	CCTTTTTCTGTGGCAAAATCGTCTCTGTACATGT	GTGTACATTTAAGAAAGGAAAGATGTT		2528

QY 2163 AAGATATGTGGCTGTGGGTACACAGGGTGCTGACGGGTAATATTTTGAATAA 2222
DB 2529 AAGATATGTGGCTGTGGGTACACAGGGTGCTGACGGGTAATATTTTGAATAA 2588
QY 2223 TATATCAAACTCACTAATCTCAATTTTAAATCAATTAATTTTCTTTT 2282
DB 2589 TATATCAAACTCACTAATCTCAATTTTAAATCAATTAATTTTCTTTT 2648
QY 2283 AAGAGAAAGCAGCTTTTCTAGACTTTAAAGAAATTAAGCTTTGGAGGCTCACGGTG 2342
DB 2649 AAGAGAAAGCAGCTTTTCTAGACTTTAAAGAAATTAAGCTTTGGAGGCTCACGGTG 2708
QY 2343 TAGAGAGAGCTTTAGAGGCCACCCGACAAATTTACCCAGAGGGAATCTGTGGAG 2402
DB 2709 TAGAGAGAGCTTTAGAGGCCACCCGACAAATTTACCCAGAGGGAATCTGTGGAG 2768
QY 2403 GACACTCAGCGAGTTCTGAGTACCTGTGTATGTCAACAGAGGAAATCCGTCTCTG 2462
DB 2769 GACACTCAGCGAGTTCTGAGTACCTGTGTATGTCAACAGAGGAAATCCGTCTCTG 2828
QY 2463 AAGAGAAAGCAGCTTTCTAGACTTTAAAGAAATTAAGCTTTGGAGGCTCACGGTG 2522
DB 2829 AAGAGAAAGCAGCTTTCTAGACTTTAAAGAAATTAAGCTTTGGAGGCTCACGGTG 2888
QY 2523 CACAGGTTTAACTGT 2582
DB 2889 CACAGGTTTAACTGT 2948
QY 2583 TCTGTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 2642
DB 2949 TCTGTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 3008
QY 2643 CTCATCTCAATCCGTATCTTACGACCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 2702
DB 3009 CTCATCTCAATCCGTATCTTACGACCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 3067
QY 2703 ATCAATCAGCAAAAGGAGCAAAAGGAGCAACCAAGCTCCATCACTTACCTTGG 2762
DB 3068 ATCAATCAGCAAAAGGAGCAAAAGGAGCAACCAAGCTCCATCACTTACCTTGG 3127
QY 2763 TTACTCAAAAGCAAGATCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2822
DB 3128 TTACTCAAAAGCAAGATCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3187
QY 2823 AACCCACCAACCAACCAATCAATCAACCAACCAACCAACCAATCAATCAATGATAT 2882
DB 3188 AACCCACCAACCAACCAATCAATCAACCAACCAACCAACCAATCAATCAATGATAT 3247
QY 2883 TTGTCTTTTGTGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2942
DB 3248 TTGTCTTTTGTGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3307
QY 2943 T-AAAAAAAAATGTGAGGAAAGTAAATTTTACCAAGTTGTGTGTGTGTGTGTGT 3001
DB 3308 TAAAAAAAAAATGTGAGGAAAGTAAATTTTACCAAGTTGTGTGTGTGTGTGTGT 3367
QY 3002 AATTCACAGTTTGT 3061
DB 3368 AATTCACAGTTTGT 3427
QY 3062 CTCAGCAGCTTGT 3121
DB 3428 CTCAGCAGCTTGT 3487
QY 3122 TTTTATTTTGT 3181
DB 3488 TTTTATTTTGT 3547
QY 3182 AGAAGGAGAGCTTGT 3241
DB 3548 AGAAGGAGAGCTTGT 3607

QY 3242 CNAATAGGCTTTTGTGACAAACCTTGAAATGTTATTT 3283
DB 3608 CTATATGGCTTTTGTGACAAACCTTGAAATGTTATTT 3649
RESULT 3
ABS76442
ID ABS76442 standard, cDNA, 3667 BP.
XX ABS76442;
AC
XX
XX
DT 11-DEC-2002 (first entry)
XX
XX
DE cDNA encoding human ovarian cancer marker M452.
XX
XX
KW Human; ovarian cancer; marker; cancer; familial history; brain disorder;
KW central nervous system disorder; bacterial meningitis; viral meningitis;
KW Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus;
KW brain herniation; inflammation; encephalitis; testicular disorder;
KW nonderculous granulomatous orchitis; connective tissue disorder;
KW heart disorder; ischaemic heart disease; atherosclerosis; neoplasm;
KW histological type; carcinogenic; ovarian cancer marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200271928-A2.
XX
PD 19-SEP-2002.
XX
PF 14-MAR-2002; 2002WO-US007826.
XX
PR 14-MAR-2001; 2001US-0276025P.
PR 14-MAR-2001; 2001US-0276026P.
PR 10-AUG-2001; 2001US-0311732P.
PR 19-SEP-2001; 2001US-0323580P.
PR 26-SEP-2001; 2001US-0324967P.
PR 26-SEP-2001; 2001US-0325102P.
PR 26-SEP-2001; 2001US-0325149P.
XX
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Monahan JR, Gannavarapu M, Hoersch S, Kamatkar S, Kovatis SG;
PI Meyers RE, Morrisey ME, Olandt PJ, Sen A, Vleidy FO, Mills GB;
PI Baet RC, Lu K, Schmandt RE, Zhao X, Ghatt K;
XX
XX
DR P-PSDB; ABG96346.
DR
XX
PT Assessing whether a patient is afflicted with ovarian cancer, useful in
PT assessing the stage or progression of the disease, comprises comparing
PT the expression level of a cancer marker in a sample from a patient and
PT from a non cancer patient.
XX
XX
PS Disclosure; Page 262-263; 481pp; English.
XX
XX
CC The present invention relates to a new method for assessing whether a
CC patient is afflicted with ovarian cancer. The method involves comparing
CC the expression level of a marker in a patient sample and the normal level
CC of expression of the marker in a control non-ovarian cancer sample, where
CC the marker is selected from 363 cancer markers described in the
CC specification. The method of the invention is useful in diagnosing or
CC characterizing cancer, in detecting the presence of cancer as early as
CC possible, and the recurrence of ovarian cancer. The method may also be of
CC particular use with patients having an enhanced risk of developing
CC ovarian cancer (e.g. patients having a familial history of ovarian
CC cancer). The cancer markers may be used in the management and treatment
CC of e.g. brain and central nervous system disorders (e.g. bacterial and
CC viral meningitis, Alzheimer's disease or Parkinson's disease), brain
CC disorders (e.g. cerebral oedema, hydrocephalus or brain herniations),
CC inflammations (e.g. bacterial or viral meningitis or encephalitis),
CC testicular disorders (e.g. nonderculous granulomatous orchitis),
CC connective tissue disorders, or heart disorders (e.g. ischaemic heart
CC disease or atherosclerosis). The compositions and methods may also be

used in assessing the histological type of neoplasm associated with ovarian cancer, monitoring the progression of ovarian cancer, determining whether ovarian cancer has metastasized or is likely to metastasize, selecting a composition for inhibiting ovarian cancer, assessing the ovarian carcinogenic potential of a compound, or inhibiting ovarian cancer or at risk of developing ovarian cancer. The present nucleic acid sequence encodes one of the ovarian cancer markers described in the invention

Sequence 3667 BP; 1081 A; 882 C; 846 G; 858 T; 0 U; 0 Other;

Query Match 99.1%; Score 3253.8; DB 6; Length 3667;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 3277; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

3 CAGCGAGAGAGCGAGAGAGCGCGGGTACCGGGCCGGGGAGACCGGGCTCTCGGGGA 62
Db CAACGAGAGAGCGAGAGCGCGGGTACCGGGCCGGGGAGACCGGGCTCTCGGGGA 428
QY GAGAGGATGATGAACAAGCTTTACATGGGAACTGAGCCCGCGCTCAAGCGGAGCA 122
Db GAGAGGATGATGAACAAGCTTTACATGGGAACTGAGCCCGCGCTCAAGCGGAGCA 488
QY GAGAGGATGATGAACAAGCTTTACATGGGAACTGAGCCCGCGCTCAAGCGGAGCA 182
Db CTTCCGGAGCTCTTTGGGGACAGAAAGCTGCCCTGGCGGACAGGTCCTGTGAAGTC 548
QY CGGCTACGCTTGTGGAGCTACCCCGACAGAACTGGGCAATCGGGCAATGAGAGCTT 608
Db CGGCTACGCTTGTGGAGCTACCCCGACAGAACTGGGCAATCGGGCAATGAGAGCTT 243
QY CTCGGGTAAAGTGAATTCATGGGAAATCATGGAATGATTAATCTAGTCTTAAAAA 302
Db CTCGGGTAAAGTGAATTCATGGGAAATCATGGAATGATTAATCTAGTCTTAAAAA 609
QY CTCGGGTAAAGTGAATTCATGGGAAATCATGGAATGATTAATCTAGTCTTAAAAA 668
Db CTCGGGTAAAGTGAATTCATGGGAAATCATGGAATGATTAATCTAGTCTTAAAAA 303
QY GCTAAAGAGCAGGAAATTCAGATTCGAAACATCCCTCTCACTGACGTGGAGGTGT 362
Db GCTAAAGAGCAGGAAATTCAGATTCGAAACATCCCTCTCACTGACGTGGAGGTGT 669
QY GCTAAAGAGCAGGAAATTCAGATTCGAAACATCCCTCTCACTGACGTGGAGGTGT 728
Db GCTAAAGAGCAGGAAATTCAGATTCGAAACATCCCTCTCACTGACGTGGAGGTGT 363
QY GGATGATCTTTGGCTCAATATGGGACAGTGGAGATGTGGAACAAATCAACAGAC 422
Db GGATGATCTTTGGCTCAATATGGGACAGTGGAGATGTGGAACAAATCAACAGAC 729
QY GGATGATCTTTGGCTCAATATGGGACAGTGGAGATGTGGAACAAATCAACAGAC 788
Db GGATGATCTTTGGCTCAATATGGGACAGTGGAGATGTGGAACAAATCAACAGAC 423
QY AGAAACCGCGCTGTCAAGCTCAATATGCAACAGAAAGAGCAAAATATGCAATGGA 482
Db AGAAACCGCGCTGTCAAGCTCAATATGCAACAGAAAGAGCAAAATATGCAATGGA 789
QY AGAAACCGCGCTGTCAAGCTCAATATGCAACAGAAAGAGCAAAATATGCAATGGA 848
Db AGAAACCGCGCTGTCAAGCTCAATATGCAACAGAAAGAGCAAAATATGCAATGGA 483
QY GAAGCTAAGCGGCAATCAGTTTGAAGACTACTCTTCAAGATTTCTAATCCCGATGA 542
Db GAAGCTAAGCGGCAATCAGTTTGAAGACTACTCTTCAAGATTTCTAATCCCGATGA 849
QY GAAGCTAAGCGGCAATCAGTTTGAAGACTACTCTTCAAGATTTCTAATCCCGATGA 908
Db GAAGCTAAGCGGCAATCAGTTTGAAGACTACTCTTCAAGATTTCTAATCCCGATGA 543
QY AGAGGTGAGTCTCCCTTCCGCTCCAGGAGCCAGCTGGGGAGCACTCTCCGGGGA 602
Db AGAGGTGAGTCTCCCTTCCGCTCCAGGAGCCAGCTGGGGAGCACTCTCCGGGGA 909
QY AGAGGTGAGTCTCCCTTCCGCTCCAGGAGCCAGCTGGGGAGCACTCTCCGGGGA 968
Db AGAGGTGAGTCTCCCTTCCGCTCCAGGAGCCAGCTGGGGAGCACTCTCCGGGGA 603
QY GCAAGGCAAGCGCCCTGGGGGCACTTCTCAAGCGCAGACAGATGATTTCCCGTGCAGAT 662
Db GCAAGGCAAGCGCCCTGGGGGCACTTCTCAAGCGCAGACAGATGATTTCCCGTGCAGAT 969
QY GCAAGGCAAGCGCCCTGGGGGCACTTCTCAAGCGCAGACAGATGATTTCCCGTGCAGAT 1028
Db GCAAGGCAAGCGCCCTGGGGGCACTTCTCAAGCGCAGACAGATGATTTCCCGTGCAGAT 663
QY CTTGTCCTCCACCAAGTTTGTGTGCTATCATGGAAGAGGCTTGAACATTAAGAA 722
Db CTTGTCCTCCACCAAGTTTGTGTGCTATCATGGAAGAGGCTTGAACATTAAGAA 1029
QY CTTGTCCTCCACCAAGTTTGTGTGCTATCATGGAAGAGGCTTGAACATTAAGAA 782
Db CTTGTCCTCCACCAAGTTTGTGTGCTATCATGGAAGAGGCTTGAACATTAAGAA 723
QY CATCTACTAAGACCAAGTCCCGGGTGAATTCATGGAAGAAAGAACTCTGAGGCTGC 1148
Db CATCTACTAAGACCAAGTCCCGGGTGAATTCATGGAAGAAAGAACTCTGAGGCTGC 1089
QY CATCTACTAAGACCAAGTCCCGGGTGAATTCATGGAAGAAAGAACTCTGAGGCTGC 842
Db CATCTACTAAGACCAAGTCCCGGGTGAATTCATGGAAGAAAGAACTCTGAGGCTGC 783
QY AGAAGACCTGTACCAATTCATGCAACCCAGAGGGGACTTCTGAAGATTCGCGAGAT 1208
Db AGAAGACCTGTACCAATTCATGCAACCCAGAGGGGACTTCTGAAGATTCGCGAGAT 1149
QY AGAAGACCTGTACCAATTCATGCAACCCAGAGGGGACTTCTGAAGATTCGCGAGAT 843
Db AGAAGACCTGTACCAATTCATGCAACCCAGAGGGGACTTCTGAAGATTCGCGAGAT

Db TCTTGAATCATGAGAAAGAGCAGATGAGACCAAACTAGCCGAGAGATTCCTGTGA 1268
QY AATCTGGCACACATGAGCTTGGTGAAGACTGATTTGAAAGAGGACAGAAATTTGAA 962
Db AATCTGGCACACATGAGCTTGGTGAAGACTGATTTGAAAGAGGACAGAAATTTGAA 1269
QY AATCTGGCACACATGAGCTTGGTGAAGACTGATTTGAAAGAGGACAGAAATTTGAA 1328
Db AATCTGGCACACATGAGCTTGGTGAAGACTGATTTGAAAGAGGACAGAAATTTGAA 963
QY GAAATTTGAACATGAAACAGGAGCCAGATTAACATCTCATCTTTGGCAGATTTGAGAGT 1022
Db GAAATTTGAACATGAAACAGGAGCCAGATTAACATCTCATCTTTGGCAGATTTGAGAGT 1329
QY GAAATTTGAACATGAAACAGGAGCCAGATTAACATCTCATCTTTGGCAGATTTGAGAGT 1082
Db GAAATTTGAACATGAAACAGGAGCCAGATTAACATCTCATCTTTGGCAGATTTGAGAGT 1023
QY AATCAACCCGGAAAGAACATCATCTGTAAGGGGACAGTTGAGGCTGTGCCAGTCTGA 1448
Db AATCAACCCGGAAAGAACATCATCTGTAAGGGGACAGTTGAGGCTGTGCCAGTCTGA 1389
QY AATCAACCCGGAAAGAACATCATCTGTAAGGGGACAGTTGAGGCTGTGCCAGTCTGA 1508
Db AATCAACCCGGAAAGAACATCATCTGTAAGGGGACAGTTGAGGCTGTGCCAGTCTGA 1449
QY GATAGAGATTATGAAAGAGCTGCTGAGAGCTTTGAAATATATGCTGCTTTAAACAC 1508
Db GATAGAGATTATGAAAGAGCTGCTGAGAGCTTTGAAATATATGCTGCTTTAAACAC 1143
QY CCACTCCGGATTCCTTCTCAGGCTGTACCCCATCAACAGATTGGCCGGTCCGGCATCA 1202
Db CCACTCCGGATTCCTTCTCAGGCTGTACCCCATCAACAGATTGGCCGGTCCGGCATCA 1509
QY CCACTCCGGATTCCTTCTCAGGCTGTACCCCATCAACAGATTGGCCGGTCCGGCATCA 1203
Db CCACTCCGGATTCCTTCTCAGGCTGTACCCCATCAACAGATTGGCCGGTCCGGCATCA 1569
QY TCACTCTTATCCAGAGCAGAGATTTGTAATCTTTCAATCCCAACCGAGCTGTGGCGC 1628
Db TCACTCTTATCCAGAGCAGAGATTTGTAATCTTTCAATCCCAACCGAGCTGTGGCGC 1263
QY CATCATCGGAGAAAGAGGGGACACATCAACAGCTGTGAGATTTGCCGGAGCTCTAT 1322
Db CATCATCGGAGAAAGAGGGGACACATCAACAGCTGTGAGATTTGCCGGAGCTCTAT 1629
QY CATCATCGGAGAAAGAGGGGACACATCAACAGCTGTGAGATTTGCCGGAGCTCTAT 1323
Db CATCATCGGAGAAAGAGGGGACACATCAACAGCTGTGAGATTTGCCGGAGCTCTAT 1689
QY CAAGATTTGCCCTGTCCGAGAGGCTCAGACCTCAGCGAAAGATGTGATATACCGGGGC 1748
Db CAAGATTTGCCCTGTCCGAGAGGCTCAGACCTCAGCGAAAGATGTGATATACCGGGGC 1383
QY ACGGAGAGCCAGTTCAAGGCCACAGGACGATCTTTGGGAAATCTGAAAGAGAAACTT 1442
Db ACGGAGAGCCAGTTCAAGGCCACAGGACGATCTTTGGGAAATCTGAAAGAGAAACTT 1749
QY ACGGAGAGCCAGTTCAAGGCCACAGGACGATCTTTGGGAAATCTGAAAGAGAAACTT 1443
Db ACGGAGAGCCAGTTCAAGGCCACAGGACGATCTTTGGGAAATCTGAAAGAGAAACTT 1809
QY CTTTAAACCCCAAGAGAGTGAAGCTGGAAGCGCATATCAAGATGCCCTCTTCCACAGC 1868
Db CTTTAAACCCCAAGAGAGTGAAGCTGGAAGCGCATATCAAGATGCCCTCTTCCACAGC 1503
QY TGGCCGGGTGATTTGGGAAAGTGGGCAAGACCGGAAACGATGCAAGACTTAAACAGTGC 1562
Db TGGCCGGGTGATTTGGGAAAGTGGGCAAGACCGGAAACGATGCAAGACTTAAACAGTGC 1869
QY TGGCCGGGTGATTTGGGAAAGTGGGCAAGACCGGAAACGATGCAAGACTTAAACAGTGC 1563
Db TGGCCGGGTGATTTGGGAAAGTGGGCAAGACCGGAAACGATGCAAGACTTAAACAGTGC 1929
QY AGAAGTCATGCTGCTGTGACCAAGCGCAGATGAAATGAGAAAGTGAATGCTCAGAT 1622
Db AGAAGTCATGCTGCTGTGACCAAGCGCAGATGAAATGAGAAAGTGAATGCTCAGAT 1929
QY AGAAGTCATGCTGCTGTGACCAAGCGCAGATGAAATGAGAAAGTGAATGCTCAGAT 1623
Db AGAAGTCATGCTGCTGTGACCAAGCGCAGATGAAATGAGAAAGTGAATGCTCAGAT 1989
QY TATCGGCACTTCTTGTGCTAGCCAGACTGACAGCGCAGATGAGGAAATTTGACACA 2048
Db TATCGGCACTTCTTGTGCTAGCCAGACTGACAGCGCAGATGAGGAAATTTGACACA 1683
QY GGTGAAGCAGCAGAGCAAGAAATACCTCAAGGAGTGCCTCAAGGCGCAGAAATGAGG 1742
Db GGTGAAGCAGCAGAGCAAGAAATACCTCAAGGAGTGCCTCAAGGCGCAGAAATGAGG 2049
QY GGTGAAGCAGCAGAGCAAGAAATACCTCAAGGAGTGCCTCAAGGCGCAGAAATGAGG 1743
Db GGTGAAGCAGCAGAGCAAGAAATACCTCAAGGAGTGCCTCAAGGCGCAGAAATGAGG 2109
QY CTTCCACAGGCAACAGCAAAACAGATGATGATGACCTTCCAAACCTGACAGATG 1802
Db CTTCCACAGGCAACAGCAAAACAGATGATGATGACCTTCCAAACCTGACAGATG 1803
QY AGACCAAAAGCAGACGACGATGCGGAGGCAAAACCAAAAGCATCTGAGAAATGAGAGT 1862
Db AGACCAAAAGCAGACGACGATGCGGAGGCAAAACCAAAAGCATCTGAGAAATGAGAGT 2169
QY AGACCAAAAGCAGACGACGATGCGGAGGCAAAACCAAAAGCATCTGAGAAATGAGAGT 1863
Db AGACCAAAAGCAGACGACGATGCGGAGGCAAAACCAAAAGCATCTGAGAAATGAGAGT 1229
QY CTGCGAGGCGGCGCAGAGACTGTGCGAGGCTTGAAGACCCAGGGGCGAGAGAGGCGC 2288
Db CTGCGAGGCGGCGCAGAGACTGTGCGAGGCTTGAAGACCCAGGGGCGAGAGAGGCGC 1923
QY GGGGAGGTGAGCCAGGTTTGGCAGAAACCAACGAGCCCGCTTCCGCTCCAGGCTT 1982
Db GGGGAGGTGAGCCAGGTTTGGCAGAAACCAACGAGCCCGCTTCCGCTCCAGGCTT 2289
QY GGGGAGGTGAGCCAGGTTTGGCAGAAACCAACGAGCCCGCTTCCGCTCCAGGCTT 2348
Db GGGGAGGTGAGCCAGGTTTGGCAGAAACCAACGAGCCCGCTTCCGCTCCAGGCTT

```
QY 1983 CTGACGGCTTCAGCCATCCACTTCACATCCATCGGATCTCTCCGAACTCCACGAGC 2042
Db 2349 CTGACGGCTTCAGCCATCCACTTCACATCCATCGGATCTCTCCGAACTCCACGAGC 2408
QY 2043 CTATCCCTTTTGAAGTGAATCAATAGTGAACGTTTCAGAACCCAGCAAAATGACAC 2102
Db 2409 CTATCCCTTTTGAAGTGAATCAATAGTGAACGTTTCAGAACCCAGCAAAATGACAC 2468
QY 2103 CTTTTTTCTGTGGCAATGTCTCTGTACATGTGTGTACATATTTGAAAAGGGAAGTGT 2162
Db 2469 CTTTTTTCTGTGGCAATGTCTCTGTACATGTGTGTACATATTTGAAAAGGGAAGTGT 2528
QY 2163 AAGATATGTGCTGTGGGTATACAGAGGTGCTGAGGGTAAATATTTATTTAGAAATTA 2222
Db 2529 AAGATATGTGCTGTGGGTATACAGAGGTGCTGAGGGTAAATATTTATTTAGAAATTA 2588
QY 2223 TATATCAATAATCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 2282
Db 2589 TATATCAATAATCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 2648
QY 2283 AAAAGAAAGAGAGGCTTTTCTAGACTTTTAAAGAAATTAAGTCTTTGGAGGTCTCA 2342
Db 2649 AAAAGAAAGAGAGGCTTTTCTAGACTTTTAAAGAAATTAAGTCTTTGGAGGTCTCA 2708
QY 2343 TAGAGAGAGGCTTTGAGGCAACCGGCAAAATTCACGAGGGAATTCGTGGAGAG 2402
Db 2709 TAGAGAGAGGCTTTGAGGCAACCGGCAAAATTCACGAGGGAATTCGTGGAGAG 2768
QY 2403 GACACTCAAGGAGTTCTGTGATTCACCTGTGTATGTCAACAGAAAGGATACCGTCTCT 2462
Db 2769 GACACTCAAGGAGTTCTGTGATTCACCTGTGTATGTCAACAGAAAGGATACCGTCTCT 2828
QY 2463 AAGAGAAACTCTGTCTCACTCTCTCATGCTGTCTAGCTATACACCATTTCTTTGCTT 2522
Db 2829 AAGAGAAACTCTGTCTCACTCTCTCATGCTGTCTAGCTATACACCATTTCTTTGCTT 2888
QY 2523 CACAGGTTTAACTGTTTTTTTGTGATCTGTATATATTTCTGTCTCTCTGTTTA 2582
Db 2889 CACAGGTTTAACTGTTTTTTTGTGATCTGTATATATTTCTGTCTCTCTGTTTA 2948
QY 2583 TCTCTCCCTCCCTCCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2642
Db 2949 TCTCTCCCTCCCTCCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3008
QY 2643 CTCATCTCAATCCCGTATCTAGCAGACCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2702
Db 3009 CTCATCTCAATCCCGTATCTAGCAGACCCCTCCCTCCCTCCCTCCCTCCCTCCCT 3067
QY 2703 ATCATCATCAAAAGGAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAG 2762
Db 3068 ATCATCATCAAAAGGAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAG 3127
QY 2763 TTAATCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAG 2822
Db 3128 TTAATCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAG 3187
QY 2823 AACCCACCAACCAACCAATCAATCAACCAACCAACCAACCAACCAACCAATCAAT 2882
Db 3188 AACCCACCAACCAACCAATCAATCAACCAACCAACCAACCAACCAACCAATCAAT 3247
QY 2883 TTTTCTTTTGTGCTTTTGTGTATTAAGCCATCAATATTCAGCAAAATGATTTCTTT 2942
Db 3248 TTTTCTTTTGTGCTTTTGTGTATTAAGCCATCAATATTCAGCAAAATGATTTCTTT 3307
QY 2943 T-AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3001
Db 3308 TAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3367
QY 3002 AATTCAGATTTTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3061
Db 3368 AATTCAGATTTTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3427
```

```
QY 3062 CTGACACCTTGTCTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTGGACATT 3121
Db 3428 CTGACACCTTGTATCTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTGGACATT 3487
QY 3122 TTTTATTTTATTAATAAATAGTGGAAAAAATPAGATATCACTGCAGCTGG 3181
Db 3488 TTTTATTTTATTAATAAATAGTGGAAAAAATPAGATATCACTGCAGCTGG 3547
QY 3182 AGAAGGTGACAGTCCAGTGTGCAACAGCTGTTCTGAATTTCTTCGCTAGCAAGAAC 3241
Db 3548 AGAAGGTGACAGTCCAGTGTGCAACAGCTGTTCTGAATTTCTTCGCTAGCAAGAAC 3607
QY 3242 CNAATATGCTCTTTTGGACAAACCTTGAAATGTTTATTT 3283
Db 3608 CTATATGCTCTTTTGGACAAACCTTGAAATGTTTATTT 3649
```

RESULT 4

AAZ36152
ID AAZ36152 standard; DNA; 3412 BP.

AAZ36152;

11-FEB-2000 (first entry)

DNA encoding cancer associated antigen KOC-3.

Cancer associated antigen; KOC-3; cancer; vaccine; CT7; ss.

Homo sapiens.

MO9954738-A1.

28-OCT-1999.

16-MAR-1999; 99WO-US005766.

17-APR-1998; 98US-00061709.

(LUDW-) LUDWIG INST CANCER RES.

Chen Y, Gure A, Zhang S, Stockert E, Jager E, Knuth A, Old LJ;

WPI; 2000-013284/01.

Nucleotides representing cancer-associated genes, used to develop products for the diagnosis, monitoring and treatment of cancers.

Claim 55; Page 41; 44pp; English.

The present sequence represents a cancer associated antigen gene designated KOC-3. The specification also describes a cancer associated antigen designated CT7. The CT7 polynucleotide was isolated from SK-MEL-37 melanoma cells. The polypeptide has some homology with MAGB-10, limited to about 210 carboxy terminal amino acids. The amino terminal of the protein has a repetitive pattern, with repeats rich in serine, proline, glutamine and leucine, and an almost invariable core of the peptide given in AAY3877. The CT7 polypeptide can be processed to peptides which provoke lysis by cytolytic T cells. The polynucleotides and polypeptides can be used for treating a cancerous condition and screening for or diagnosing cancerous conditions. The cancer associated antigens can be used as an immunogenic or vaccine composition with an adjuvant, e.g. a cytokine, a saponin, or granulocyte macrophage-colony stimulating factor (GM-CSF).

Sequence 3412 BP; 970 A; 887 C; 804 G; 750 T; 0 U; 1 Other;

Query Match 95.7%; Score 3143; DB 3; Length 3412;

Best Local Similarity 96.2%; Pred. No. 0;

Matches 3283; Conservative 0; Mismatches 0; Indels 129; Gaps 1;

1 GGCAAGCGAGAGCGCAGAGCGCGGCTACCGGCGCGGGAGCGCGGCTCTCGGG 60

Db 1 GGCACGAGAGAGGAGGAGGCGCGGATACCGGGCCGGGGAGGCCGCGCTCCGAGG 60
QY 61 AAGAGACGATGATGAAACAAGCTTTACATCGGAAACTGAGCCCGGCGCTCACCCGAC 120
Db 61 AAGAGACGATGATGAAACAAGCTTTACATCGGAAACTGAGCCCGGCGCTCACCCGAC 120
QY 121 GACCTCCGGCAGCTCTTTGGGGGACAGAGAGCTCCCTGGCGGGACAGGCTCCGTGAAG 180
Db 121 GACCTCCGGCAGCTCTTTGGGGGACAGAGAGCTCCCTGGCGGGACAGGCTCCGTGAAG 180
QY 181 TCCGCTACGCTCTGCTGAGTACCCCGACCAAGACTGGGCGCATCCGAGAC 240
Db 181 TCCGCTACGCTCTGCTGAGTACCCCGACCAAGACTGGGCGCATCCGAGAC 240
QY 241 CTCTCGGGTAAAGTGAATTTGCAATGGGAAATCATGGAAGTTGATTAATCAAGTCTTAA 300
Db 241 CTCTCGGGTAAAGTGAATTTGCAATGGGAAATCATGGAAGTTGATTAATCAAGTCTTAA 300
QY 301 AAGCTAAGGAGAGGAAATTCAGATTGCAACATCCCTCCTCAGCTGAGTGGAGTG 360
Db 301 AAGCTAAGGAGAGGAAATTCAGATTGCAACATCCCTCCTCAGCTGAGTGGAGTG 360
QY 361 TTGATGATGACTTTTGGCTCAATATGAGGACAGTGAATGTGGAACAAGTCAACAGAC 420
Db 361 TTGATGATGACTTTTGGCTCAATATGAGGACAGTGAATGTGGAACAAGTCAACAGAC 420
QY 421 ACAGAAACCGCGCTTGTCAACGTCAATATGCAACAGAGAAAGACCAAAATAGCCATG 480
Db 421 ACAGAAACCGCGCTTGTCAACGTCAATATGCAACAGAGAAAGACCAAAATAGCCATG 480
QY 481 GAGAAGCTAAGCGGAGTCAAGTTTGAAGCTACTCTTCAAGATTTCTTCAATCCCGAT 540
Db 481 GAGAAGCTAAGCGGAGTCAAGTTTGAAGCTACTCTTCAAGATTTCTTCAATCCCGAT 540
QY 541 GAAGAGTGAAGTCCCTTTCGCCCCCTCAGCGAGCCGAGGTGGGACCACTTTCCCG 600
Db 541 GAAGAGTGAAGTCCCTTTCGCCCCCTCAGCGAGCCGAGGTGGGACCACTTTCCCG 600
QY 601 GAGCAAGGCTACGCCCCCTGGGGGCACTTCTCAAGGCGACAGATTAATTTCCCGCTGG 660
Db 601 GAGCAAGGCTACGCCCCCTGGGGGCACTTCTCAAGGCGACAGATTAATTTCCCGCTGG 660
QY 661 ATCTGCTCCCAACCCAGTTTGTGGTCCATCATCGAAAGAGGCTTGAACATTAAG 720
Db 661 ATCTGCTCCCAACCCAGTTTGTGGTCCATCATCGAAAGAGGCTTGAACATTAAG 720
QY 721 AACATCACTAAGACAGCCAGTCCCGGGTAGATATCCATAGAAAGAGAACTCTGAGCT 780
Db 721 AACATCACTAAGACAGCCAGTCCCGGGTAGATATCCATAGAAAGAGAACTCTGAGCT 780
QY 781 GCGAGAGGCTGTCACTATCCATGCAACCCGAGAGGGGACTTCTGAAGATGCCGATG 840
Db 781 GCGAGAGGCTGTCACTATCCATGCAACCCGAGAGGGGACTTCTGAAGATGCCGATG 840
QY 841 ATTCTTGAATCATGCGAAAGAGGAGAGTGAAGCAAACTAGCGAAGAGATTCCTG 900
Db 841 ATTCTTGAATCATGCGAAAGAGGAGAGTGAAGCAAACTAGCGAAGAGATTCCTG 900
QY 901 AAAATCTTGGCACAATGCTGCTGGTGGAAAGTGAATGGAAAGAGGAGGAGAAATTTG 960
Db 901 AAAATCTTGGCACAATGCTGCTGGTGGAAAGTGAATGGAAAGAGGAGGAGAAATTTG 960
QY 961 AAGAAATTTGAACATGAAACAGGGACCAAGATTAACATTCATTTTGGCAGGATTTAGC 1020
Db 961 AAGAAATTTGAACATGAAACAGGGACCAAGATTAACATTCATTTTGGCAGGATTTAGC 1020
QY 1021 ATATACACCGGAGAAAGACCATCATGTGAAGGCAAGTTGAGGCGCTGTGCCAGTCT 1080
Db 1021 ATATACACCGGAGAAAGACCATCATGTGAAGGCAAGTTGAGGCGCTGTGCCAGTCT 1080
QY 1081 GAGATAGAGTTATGAAGAGCTGCTGAGGCTTTGAAATGATATGCTGGCTTTAAC 1140
Db 1081 GAGATAGAGTTATGAAGAGCTGCTGAGGCTTTGAAATGATATGCTGGCTTTAAC 1140

QY 1141 ----- 1140
Db 1141 CAACAAGCCATCTGATCCAGGGTTGAACCTCAGGCACTTGGCATTTTCAACAGGA 1200
QY 1141 ----- 1140
Db 1201 CTGTCCGTGCTATCTTCAACAGAGGCCCCGAGAGTCCCCCGCTGCCCTTACAC 1260
QY 1141 -----ACCCACTCCGGATTAATTTCTCAGCTGTACCCCCATCAGCATTTGGCCGG 1191
Db 1261 CCTTCTACTACCCACTCCGGATTAATTTCTCAGCTGTACCCCCATCAGCATTTGGCCGG 1320
QY 1192 TTCCGATCATCATCTTATCTTATCAGAGCAGAGATTTGATCTTCAATCCCAACCA 1251
Db 1321 TTCCGATCATCATCTTATCTTATCAGAGCAGAGATTTGATCTTCAATCCCAACCA 1380
QY 1252 GCTGTGGGCGCATCATCGGGAAGAAAGGGGCAACATCAACAGCTGGCGAGATTGGCC 1311
Db 1381 GCTGTGGGCGCATCATCGGGAAGAAAGGGGCAACATCAACAGCTGGCGAGATTGGCC 1440
QY 1312 GGAAGCTCTATCAAGTTGCCCCCTGGGAAAGGCCAGAGCTCAGCGAAAGATGTATC 1371
Db 1441 GGAAGCTCTATCAAGTTGCCCCCTGGGAAAGGCCAGAGCTCAGCGAAAGATGTATC 1500
QY 1372 ATCACCGGGCCACCGGAAACCCAGTTCAAGGCCAGGACAGATCTTTGGAAATCTGAAA 1431
Db 1501 ATCACCGGGCCACCGGAAACCCAGTTCAAGGCCAGGACAGATCTTTGGAAATCTGAAA 1560
QY 1432 GAGAAAACTTTTAAACCCCAAGAAAGATGAACTGGAACCGCATTCAGAGTGGCCC 1491
Db 1561 GAGAAAACTTTTAAACCCCAAGAAAGATGAACTGGAACCGCATTCAGAGTGGCCC 1620
QY 1492 TCTTCCACAGCTGGCGGGGATTTGGCAAGGTGGCAAGACCTGTGAACGAATCGGAAAC 1551
Db 1621 TCTTCCACAGCTGGCGGGGATTTGGCAAGGTGGCAAGACCTGTGAACGAATCGGAAAC 1680
QY 1552 TTAAACAGTGCAGAAATCATCTGCTGTGACCAACCGCAGATGAAATGAGAAAGTG 1611
Db 1681 TTAAACAGTGCAGAAATCATCTGCTGTGACCAACCGCAGATGAAATGAGAAAGTG 1740
QY 1612 ATGTGCAAAATTTATGGGCACTTTTGTGATGCAACATGCAAGGCAAGATCAAGGAA 1671
Db 1741 ATGTGCAAAATTTATGGGCACTTTTGTGATGCAACATGCAAGGCAAGATCAAGGAA 1800
QY 1672 ATTGTAAACAGGTGAAGCAGAGCAGAGAAATACCTCAGGAGTGGCTCAAGCGC 1731
Db 1801 ATTGTAAACAGGTGAAGCAGAGCAGAGAAATACCTCAGGAGTGGCTCAAGCGC 1860
QY 1732 AGCAAGTGAAGCTCCACAGGCAACGAAACCAACGAGATGATGATGCTTCCAAAC 1791
Db 1861 AGCAAGTGAAGCTCCACAGGCAACGAAACCAACGAGATGATGATGCTTCCAAAC 1920
QY 1792 CTGACGAATGAAACCAAGCAGCAGCAGCAGATCCGGGACCAAGAGCATCTGAG 1851
Db 1921 CTGACGAATGAAACCAAGCAGCAGCAGCAGATCCGGGACCAAGAGCATCTGAG 1980
QY 1852 GAATGAAAGTCTGCGGAGCGGCAAGGAGCTGTCCGAGGCTTGAAGACCCAGGGG 1911
Db 1981 GAATGAAAGTCTGCGGAGCGGCAAGGAGCTGTCCGAGGCTTGAAGACCCAGGGG 2040
QY 1912 CGAGAGGCGGAGGAGGATCAGCAGGTTTGCAGAACCAAGAGCCCGCTCCCGCC 1971
Db 2041 CGAGAGGCGGAGGAGGATCAGCAGGTTTGCAGAACCAAGAGCCCGCTCCCGCC 2100
QY 1972 CCCCAGGCTTTCTGAGGCTTCAAGCATCACTTCAACATCCACTGGATCTCTGAA 2031
Db 2101 CCCCAGGCTTTCTGAGGCTTCAAGCATCACTTCAACATCCACTGGATCTCTGAA 2160
QY 2032 CTCCACAGACGCTATCCCTTTAGTTGAATGAATAGTGAACGTGTTCAAGCCAAGC 2091
Db 2161 CTCCACAGACGCTATCCCTTTAGTTGAATGAATAGTGAACGTGTTCAAGCCAAGC 2220

QY 2092 AAAATGACACCCCTTTTCTGTGGCAATGCTCTGTACATGTGTGACATATTAGAAA 2151
 DB 2221 AAAATGACACCCCTTTTCTGTGGCAATGCTCTGTACATGTGTGACATATTAGAAA 2280
 QY 2152 GGAAGATGTAAAGATATGTGGCTGTGGTTACACAGAGTGTCCGACGGTAATAT 2211
 DB 2281 GGAAGATGTAAAGATATGTGGCTGTGGTTACACAGAGTGTCCGACGGTAATAT 2340
 QY 2212 TTAGAAATATATATCAATTAATCACTAATCAATTTTAAATCAATTAATTTT 2271
 DB 2341 TTAGAAATATATATCAATTAATCACTAATCAATTTTAAATCAATTAATTTT 2400
 QY 2272 TTTTCTTTTAAAGAAAGAGGCTTTCTAGACTTTAAAGATTAAGCTTTGGAG 2331
 DB 2401 TTTTCTTTTAAAGAAAGAGGCTTTCTAGACTTTAAAGATTAAGCTTTGGAG 2460
 QY 2332 GTCTACGGGTGTAAGAGAGGCTTTGAGGCCACCCGCAAAATTCACCCAGAGGAAAT 2391
 DB 2461 GTCTACGGGTGTAAGAGAGGCTTTGAGGCCACCCGCAAAATTCACCCAGAGGAAAT 2520
 QY 2392 CTCTCGAGAGGACACTACGGGCACTTGTGATCACTGTGTATGTCAACGAAGGATA 2451
 DB 2521 CTCTCGAGAGGACACTACGGGCACTTGTGATCACTGTGTATGTCAACGAAGGATA 2580
 QY 2452 CCGTCTCTTGAAGAGAACTGTGTCACTCTCAATGCTGTAGCTCATACACCCATT 2511
 DB 2581 CCGTCTCTTGAAGAGAACTGTGTCACTCTCAATGCTGTAGCTCATACACCCATT 2640
 QY 2512 TCTCTTTGCTTCAACAGTTTAAACGTGTTTGTGATATGTCTATATATTTCTGTCT 2571
 DB 2641 TCTCTTTGCTTCAACAGTTTAAACGTGTTTGTGATATGTCTATATATTTCTGTCT 2700
 QY 2572 CTCTCTGTTTATCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2631
 DB 2701 CTCTCTGTTTATCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2760
 QY 2632 TTTCTCATCCCTCATCTCAATCCGATCTACGACCCCTCCCTCCCTCCCTCCCTCCCT 2691
 DB 2761 TTTCTCATCCCTCATCTCAATCCGATCTACGACCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2820
 QY 2692 GTGCTCGAGTATCACTACACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2751
 DB 2821 GTGCTCGAGTATCACTACACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2880
 QY 2752 CTACACTGTGTACTCAAAAGAAAGAAAGTCAATGATCTGTGCTTGTGCTTGTGAG 2811
 DB 2881 CTACACTGTGTACTCAAAAGAAAGAAAGTCAATGATCTGTGCTTGTGCTTGTGAG 2940
 QY 2812 AGGAAACAGGAAACCAACCAACCAATCAACCAACCAAGAAAGAAAGAAAGAAAG 2871
 DB 2941 AGGAAACAGGAAACCAACCAACCAATCAACCAACCAAGAAAGAAAGAAAGAAAG 3000
 QY 2872 AAAGAAAGTATTTTGTCTTTTGTGATTTGTGTATTAAGCCATATATTCAGAAATG 2931
 DB 3001 AAAGAAAGTATTTTGTCTTTTGTGATTTGTGTATTAAGCCATATATTCAGAAATG 3060
 QY 2932 ATTCTCTTCTTAAAGAAAGAAAGTGAAGAAATTAAGCAAGTGTGGCC 2991
 DB 3061 ATTCTCTTCTTAAAGAAAGAAAGTGAAGAAATTAAGCAAGTGTGGCC 3120
 QY 2992 CAGGGCGTTAATTCACAGATTTTAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 3051
 DB 3121 CAGGGCGTTAATTCACAGATTTTAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 3180
 QY 3052 GTGTTTAACTACAGACCTGTCTGTGTCTTCCCTTAAGATTTTGAAGCTGATAG 3111
 DB 3181 GTGTTTAACTACAGACCTGTCTGTGTCTTCCCTTAAGATTTTGAAGCTGATAG 3240
 QY 3112 TTGAGACATTTTAAATTTTAAATTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 3171
 DB 3241 TTGAGACATTTTAAATTTTAAATTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 3300
 QY 3172 GCCAGCCTGAGAGAGGTGACAGTCCAGTGTGCAAGCTGTGTAATGTCTCCGCT 3231

DB 3301 GCCAGCCTGAGAGAGTATGACACTCCAAAGTGTGCAACAGCTGTTGAAATGTCTCCGCT 3360
 QY 3232 AGCCAGAACCNATATGACCTTTTGTGACAAACCTTGAAGAAATGTTAATTT 3283
 DB 3361 AGCCAGAACCNATATGACCTTTTGTGACAAACCTTGAAGAAATGTTAATTT 3412
 RESULT 5
 AAS26150
 ID AAS26150 standard, cDNA, 3694 BP.
 XX
 AC AAS26150;
 XX
 DT 07-NOV-2001 (first entry)
 XX
 DE Human cDNA encoding a novel secreted protein, Seq ID 329.
 KW
 KW Human; immunosuppressive; antiarthritic; ss; antineumatic; cyostatic;
 KW cardiatic; vasotropic; cerebroprotective; neurotropic; neuroprotective;
 KW antibacterial; virocidic; fungicide; ophthalmological; vulnery;
 KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
 KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 KW cerebral ischaemia; angiogenesis; nervous system disorder;
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;
 KW wound healing; epithelial cell proliferation; skin ageing; food additive;
 KW preservative; antiproliferative.
 XX
 OS Homo sapiens.
 XX
 PN WO20015322-A2.
 XX
 PD 02-AUG-2001.
 XX
 PE 17-JAN-2001; 2001WO-US001341.
 XX
 PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205151P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226688P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230433P.
PR 06-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-023397P.
PR 14-SEP-2000; 2000US-023398P.
PR 14-SEP-2000; 2000US-023399P.
PR 14-SEP-2000; 2000US-0233400P.
PR 14-SEP-2000; 2000US-023401P.
PR 14-SEP-2000; 2000US-023063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-023497P.
PR 25-SEP-2000; 2000US-023498P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241866P.
PR 01-NOV-2000; 2000US-024617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.

PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-48783/53.
P-PSDB; AAU16163.

PT New nucleic acid molecules encoding 461 human secreted proteins for
diagnosing, preventing, treating or ameliorating medical conditions and
used as food additives or preservatives.

XX Claim 1; SEQ ID NO 329; 980bp; English.

XX The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Antibodies to the proteins can also be used in
CC alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC and many other disorders listed in the specification. The polypeptides
CC can also be used to aid wound healing and epithelial cell proliferation,
CC to prevent skin aging due to sunburn, to maintain organ before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues or preservative to increase or decrease storage
CC as a food additive or preservative to increase or decrease storage
CC capability, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence encodes a novel secreted protein of the invention. Note: The
CC sequence data for this patent did not form part of the printed

Query Match 93.0%; Score 3052.6; DB 4; Length 3694;

Best Local Similarity 96.0%; Pred. No. 0; Indels 132; Gaps 3;
Matches 3214; Conservative 1; Mismatches 1;

QY 68 GGATGATGACCAAGCTTACATCGGAACCTGAGCCCGCGCTACCGCGGACGACTCC 127
DB 91 GGATGATGACCAAGCTTACATCGGAACCTGAGCCCGCGCTACCGCGGACGACTCC 150

```
OY 128 GGCAGCTCTTTGGGGACAGAAAGCTGCCCTGGCGGGACAGGTCTCTGAAGTCCGCT 187
Db 151 GGCAGCTCTTTGGGGACAGAAAGCTGCCCTGGCGGGACAGGTCTCTGAAGTCCGCT 210
OY 188 AGGCTTCTGTGACATACCCCGACACAGAACTGGGCACTCCGCGCATGAGACCTCTCGG 247
Db 211 AGGCTTCTGTGACATACCCCGACACAGAACTGGGCACTCCGCGCATGAGACCTCTCGG 270
OY 248 GTAAAGTGAATTGATGGGAAATCATGGAAGTTGATTACTCAGTCTTAAAAAGCTTA 307
Db 271 GTAAAGTGAATTGATGGGAAATCATGGAAGTTGATTACTCAGTCTTAAAAAGCTTA 330
OY 308 GGAGGAGGAAAAATTCAGATTGGAATCCCTCTCACTGCAAGTGGGAGGTTGGAAAG 367
Db 331 GGAGGAGGAAAAATTCAGATTGGAATCCCTCTCACTGCAAGTGGGAGGTTGGAAAG 390
OY 368 GACTTTTGGCTCAATATGAGGACAGTGAAGATGTGAAAACAACTCAACACAGACAGAA 427
Db 391 GACTTTTGGCTCAATATGAGGACAGTGAAGATGTGAAAACAACTCAACACAGACAGAA 450
OY 428 CGGCGGTGTCAAGSTCACTATGCAACAGAGAAAGCAAAATATGCAATGAGAAC 487
Db 451 CGGCGGTGTCAAGSTCACTATGCAACAGAGAAAGCAAAATATGCAATGAGAAC 510
OY 488 TAAAGGCGCATCAGTTTGAAGACTACTCTTCAAGATTTCCTACATCCCGATGAAGG 547
Db 511 TAAAGGCGCATCAGTTTGAAGACTACTCTTCAAGATTTCCTACATCCCGATGAAGG 570
OY 548 TGAAGTCCCTTTCGCCCCCTTCAAGCGAGCCAGCGTGGGAGCACTCTTCCGCGAGCAAG 607
Db 571 TGAAGTCCCTTTCGCCCCCTTCAAGCGAGCCAGCGTGGGAGCACTCTTCCGCGAGCAAG 630
OY 608 GGCAGCGCCCTGGGGGGCACTTCTCAGGCGACACAGATTGATTTCCCGCTCGGATCTGG 667
Db 631 GGCAGCGCCCTGGGGGGCACTTCTCAGGCGACACAGATTGATTTCCCGCTCGGATCTGG 690
OY 668 TCCCCACCAGTTTGTGTGTCATCATCGGAAAGAGGGCTTGAACATAAAGAACATCA 727
Db 691 TCCCCACCAGTTTGTGTGTCATCATCGGAAAGAGGGCTTGAACATAAAGAACATCA 750
OY 728 CTAAAGCAGACCCAGTCCCGGGTAGATATCATAGAAAAGAGAACTCTGGAAGCTGACAGA 787
Db 751 CTAAAGCAGACCCAGTCCCGGGTAGATATCATAGAAAAGAGAACTCTGGAAGCTGACAGA 810
OY 788 AGCGTGCACATCATATGCCAAGCCAGAGGGGACTTCTGAAGCATGCCGATGATTTCTTG 847
Db 811 AGCGTGCACATCATATGCCAAGCCAGAGGGGACTTCTGAAGCATGCCGATGATTTCTTG 870
OY 848 AATCATGCAAGAAAGGACAGATGAGACCAACTAGCCGAGAGATTCCTGAAAAATCT 907
Db 871 AATCATGCAAGAAAGGACAGATGAGACCAACTAGCCGAGAGATTCCTGAAAAATCT 930
OY 908 TGGCACAACATGCTGTGGTGGAAAGCTGATTGAAAAAGAGGACAGAAATTTGAAGAAA 967
Db 931 TGGCACAACATGCTGTGGTGGAAAGCTGATTGAAAAAGAGGACAGAAATTTGAAGAAA 990
OY 968 TTGAACATGGAAGAGGAGCAAGATAAATAATCTCATCTTTGCAAGATTTGACATATACA 1027
Db 991 TTGAACATGGAAGAGGAGCAAGATAAATAATCTCATCTTTGCAAGATTTGACATATACA 1050
OY 1028 ACCCGGAAAGAACATCAGTGTGAAGGACAGTTGAGGCTGTGACAGTGTGAGATAG 1087
Db 1051 ACCCGGAAAGAACATCAGTGTGAAGGACAGTTGAGGCTGTGACAGTGTGAGATAG 1110
OY 1088 AGATTATGAAGAGCTGCTGAAGGCTTTGAAAAATGATGCTGCTGTTAAAC----- 1140
Db 1111 AGATTATGAAGAGCTGCTGAAGGCTTTGAAAAATGATGCTGCTGTTAAACAGAG 1170
OY 1141 ----- 1140
Db 1171 CCAATCTGATCCAGGCTTGAACCTCAGCGCAGCTTGCGATCTTTCAACAGAGCTGTGCG 1230
OY 1141 ----- 1140
Db 1231 TGCTATCTCACACAGAGGCCCCCGGAGCTCCCCCGGCTGCCCCCTACACACCCCTTCA 1290
OY 1141 --ACCCACTCCGGATACTTCTTCAGCCTGTACCCCCATACACAGTTTGGCCGTTCCGCG 1198
Db 1291 CTACCCACTCCGGATACTTCTTCAGCCTGTACCCCCATACACAGTTTGGCCGTTCCGCG 1350
OY 1199 ATCATCATCTTTATCTCAAGACAGAGATTGTGAATCTTTCATCCCAACCCAGGCTGTGG 1258
Db 1351 ATCATCATCTTTATCTCAAGACAGAGATTGTGAATCTTTCATCCCAACCCAGGCTGTGG 1410
OY 1259 GGGCCATCATGGGAGAAAGGGGGGACACATCAAAACACTGGCGAGATTTCCGCGAGCCT 1318
Db 1411 GGGCCATCATGGGAGAAAGGGGGGACACATCAAAACACTGGCGAGATTTCCGCGAGCCT 1470
OY 1319 CTATCAAGATTGCCCCCTCGAGAGGCCAGACGTGAGGAGAAAGATGTCATCATCACCG 1378
Db 1471 CTATCAAGATTGCCCCCTCGAGAGGCCAGACGTGAGGAGAAAGATGTCATCATCACCG 1530
OY 1379 GGCCACCGGAAAGCCCATGTTCAAGGCTCAGGAGCCGATCTTTGGGAACTGAAAGAGAAA 1438
Db 1531 GGCCACCGGAAAGCCCATGTTCAAGGCTCAGGAGCCGATCTTTGGGAACTGAAAGAGAAA 1590
OY 1439 ACTTCTTTAACCCCAAGAGAGTGAAGTGGAAAGCGCATATGAGATGCGCTCTTCCA 1498
Db 1591 ACTTCTTTAACCCCAAGAGAGTGAAGTGGAAAGCGCATATGAGATGCGCTCTTCCA 1650
OY 1499 CAGCTGGCCGGGTGATTTGGCAAAAGGTGGCAAGACCGTGAAGAACTGACAGAACTTTACA 1558
Db 1651 CAGCTGGCCGGGTGATTTGGCAAAAGGTGGCAAGACCGTGAAGAACTGACAGAACTTTACA 1710
OY 1559 GTGCAGAAATCATGCTGTGCTGTGACCAAAACGCCAGATGAATAATGAGAAATGTGCTCA 1618
Db 1711 GTGCAGAAATCATGCTGTGCTGTGACCAAAACGCCAGATGAATAATGAGAAATGTGCTCA 1770
OY 1619 GAATTTATCGGGCAGCTTCTTGTGCGCAGATGCAAGCGGCAATATGAGGAAATTTGAC 1678
Db 1771 GAATTTATCGGGCAGCTTCTTGTGCGCAGATGCAAGCGGCAATATGAGGAAATTTGAC 1830
OY 1679 AACAGTGAACACACAGAGACAGAAATACCTTCAGGAGTGTGCTTCACAGGCGACAGAGT 1738
Db 1831 AACAGTGAACACACAGAGACAGAAATACCTTCAGGAGTGTGCTTCACAGGCGACAGAGT 1890
OY 1739 GAGGCTCCACAGGACCAAGCAAAACAAACGAGTAAATGTAGCTCTTCCACACCTGACAG 1798
Db 1891 GAGGCTCCACAGGACCAAGCAAAACAAACGAGTAAATGTAGCTCTTCCACACCTGACAG 1950
OY 1799 AATGAGACCAAGCGACAGCGCCAGATCGGAGAGCAACCAAAAGCACTTGAAGAAATGAG 1858
Db 1951 AATGAGACCAAGCGACAGCGCCAGATCGGAGAGCAACCAAAAGCACTTGAAGAAATGAG 2010
OY 1859 AAGTCTCGGAGGCGGACAGGAGCTGTGCGAGGCTGTGAGAACCCACAGGGGCGGAGAG 1918
Db 2011 AAGTCTCGGAGGCGGACAGGAGCTGTGCGAGGCTGTGAGAACCCACAGGGGCGGAGAG 2070
OY 1919 GGGCGGGAAAGGTGACCGAGGTTTGGCAAGACACGAGAGCCCGCTCCGCGCCCAAG 1978
Db 2071 GGGCGGGAAAGGTGACCGAGGTTTGGCAAGACACGAGAGCCCGCTCCGCGCCCAAG 2130
OY 1979 GCTTCTGACAGGCTTCAAGCATCACTTCCACATCCATCTCGGATCTCTCTGAACTCCAC 2038
Db 2131 GCTTCTGACAGGCTTCAAGCATCACTTCCACATCCATCTCGGATCTCTCTGAACTCCAC 2190
OY 2039 GAGGCTATCCCTTTTATTTGAACCTAATAGTGAAGCTGTTCAAACCCAAAGCAAAATGC 2098
Db 2191 GAGGCTATCCCTTTTATTTGAACCTAATAGTGAAGCTGTTCAAACCCAAAGCAAAATGC 2250
OY 2099 ACACCTTTTCTGTGCAAAATGCTCTGTATCATGTGTGTATCATATTAGAAAGGAGAGA 2158
Db 2251 ACACCTTTTCTGTGCAAAATGCTCTGTATCATGTGTGTATCATATTAGAAAGGAGAGA 2310
OY 2159 TGTTAAGATATGAGGCTGTGGGTTTACAGGGGCTGTGACGGGTAAATATTTTGA 2218
```

Db 2311 TGTTAAGATATGTCGCTGTGGGTTACACAGGTCGCTGACGGGTAAATATTTAGAA 2370
Qy 2219 ATAAATATATCAAAATACCTCACTAACTCCAAATTTTAAATTAATTAATTTTCT 2278
Db 2371 ATAAATATCAAAATACCTCACTAACTCCAAATTTTAAATTAATTAATTTTCT 2430
Qy 2279 TTTTAAAGAAAGAGGCTTTCTAGACTTAAAGAAATTAAGCTTTGGAGGCTCAC 2338
Db 2431 TTTTAAAGAAAGAGGCTTTCTAGACTTAAAGAAATTAAGCTTTGGAGGCTCAC 2490
Qy 2339 GGTGTAAGAGAGGCTTTGAGGCCACCCGACAAAATTCACCAAGAGGAAATCTGTCG 2398
Db 2431 GGTGTAAGAGAGGCTTTGAGGCCACCCGACAAAATTCACCAAGAGGAAATCTGTCG 2550
Qy 2399 GAAGGACACTCAGGCGAGTTCTGATCACTGCTGATGTCACACAGAGGAAATCCGCTGC 2458
Db 2551 GAAGGACACTCAGGCGAGTTCTGATCACTGCTGATGTCACACAGAGGAAATCCGCTGC 2610
Qy 2459 CTTGAAGAGGAAAGCTGTCACCTCCATGCGCTGCTAGCTCATACCCATTTCTCTT 2518
Db 2611 CTTGAAGAGGAAAGCTGTCACCTCCATGCGCTGCTAGCTCATACCCATTTCTCTT 2670
Qy 2519 GCTTCAAGGTTTAACTGATTTTGTGACATCTGCTATTAATTTCTGCTCTCTG 2578
Db 2671 GCTTCAAGGTTTAACTGATTTTGTGACATCTGCTATTAATTTCTGCTCTCTG 2730
Qy 2579 TTTATCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTC 2638
Db 2731 TTTATCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTC 2790
Qy 2639 ATCCCTCCATCTCAATCCCGATCTCAAGCA - CCCCCCCCCCAGGCAAGCAAGCT 2696
Db 2791 ATCCCTCCATCTCAATCCCGATCTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCT 2850
Qy 2697 CTGAGTATCATCATCACAAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAG 2756
Db 2851 CTGAGTATCATCATCACAAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAG 2910
Qy 2757 ACTTGGTACTCAAAAGAAAGAGTCAATGTAATTTGCTTCTGCTTTTGGAGAGGAA 2816
Db 2911 ACTTGGTACTCAAAAGAAAGAGTCAATGTAATTTGCTTCTGCTTTTGGAGAGGAA 2970
Qy 2817 AACAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAA 2876
Db 2971 AACAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAA 3030
Qy 2877 ATGTAATTTGCTCTTTTGTGATTTTGTGTAAGCAATCAATTTCAAGAAATGATTC 2936
Db 3031 ATGTAATTTGCTCTTTTGTGATTTTGTGTAAGCAATCAATTTCAAGAAATGATTC 3090
Qy 2937 TTTCTTT -AAAAAAGAGAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGG 2995
Db 3091 TTTCTTTAAAAAAGAGAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGG 3150
Qy 2996 GCGTAAATCACAGATTTTAAAGAGAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGG 3055
Db 3151 GCGTAAATCACAGATTTTAAAGAGAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGG 3210
Qy 3056 TTTTACCTGACACCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 3115
Db 3211 TTTTACCTGACACCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 3270
Qy 3116 AGCATTTTATTTTAAATTAAGAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAG 3175
Db 3271 AGCATTTTATTTTAAATTAAGAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAG 3330
Qy 3176 GCGTGAAGAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGG 3235
Db 3331 GCGTGAAGAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGG 3390
Qy 3236 AAGAACCAATATGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 3283
Db 3391 AAGAACCAATATGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 3438

RESULT 6
ABX73491
ID ABX73491 standard; DNA; 3694 BP.
XX
AC ABX73491;
XX
DT 18-MAR-2003 (first entry)
XX
DE Human novel polynucleotide #319.
XX
KW Human; gene; ds; neural disorder; immune system disorder; renal disorder;
KW muscular disorder; respiratory disease; reproductive disorder;
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW hyperproliferative disorder; inflammatory disease; allergic reaction;
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW cardiovascular; nephrotoxic; cytostatic; antiallergic; thrombolytic;
KW haemostatic; antiarteriosclerotic.
XX
OS Homo sapiens.
XX
PN US2002132753-A1.
XX
PD 19-SEP-2002.
XX
PF 17-JAN-2001; 2001US-00764864.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 14-AUG-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225470P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0226688P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234977P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239355P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.

PR 01-NOV-2000; 2000US-0244617P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 XX
 PA (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX
 PI Rosen CA, Ruben SM, Barash SC;
 XX
 DR WPI; 2003-147444/14.
 DR P-PSDB; ABUS5231.
 DR
 XX
 PT New polypeptides and nucleic acids, useful in gene therapy for treating,
 PT inhibiting or preventing e.g. neural, immune system, muscular,
 PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
 PT renal disorders.
 XX
 PS Claim 1; SEQ ID NO 329; 402bp; English.
 XX

The invention relates to human novel polypeptides and their associated polynucleotides. The polypeptides and polynucleotides are useful in gene therapy for treating, inhibiting or preventing neural disorders, immune system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis and multiple sclerosis), muscular disorders, respiratory diseases (e.g. nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders, gastrointestinal disorders, pulmonary disorders, cardiovascular disorders (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left heart syndrome), renal disorders (e.g. acute kidney failure and end-stage renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and leukemia), inflammatory diseases (e.g. septic shock, bursitis and appendicitis), allergic reactions and conditions (e.g. asthma), blood related disorders (e.g. thrombosis, atherosclerosis and myocardial infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent human novel polynucleotides of the invention

Sequence 3694 BD; 1103 A; 934 C; 839 G; 817 T; 0 U; 1 Other;

Query Match 93.0%; Score 3052.6; DB 7; Length 3694;
 Best Local Similarity 96.0%; Pred. No. 0;

Matches 3214; Conservative 1; Mismatches 1; Indels 132; Gaps 3;

QY 68 GGATGATGAACAAGCTTTAATGAGGAACTGAGCCCGCCCTGACCGCGAGACCTCC 127
 |||
 DB 91 GGATGATGAACAAGCTTTAATGAGGAACTGAGCCCGCCCTGACCGCGAGACCTCC 150
 |||
 QY 128 GGCAGCTCTTTGGGAGACAGGAAGCTGCCCTGGCGGAGACAGTCTCTGAAGTCGGCT 187
 |||
 DB 151 GGCAGCTCTTTGGGAGACAGGAAGCTGCCCTGGCGGAGACAGTCTCTGAAGTCGGCT 210
 |||
 QY 188 ACCGCTTCCTGAGACTACCCCGACCAAGAACTGGGCGCATCGCGCCATGAGACCTCTCGG 247
 |||
 DB 211 ACCGCTTCCTGAGACTACCCCGACCAAGAACTGGGCGCATCGCGCCATGAGACCTCTCGG 270
 |||
 QY 248 GTAAGTGAATTCATGAGGAAATCATGGAAGTGAATTAATCTAGTCTTAAAAAGTTAA 307
 |||
 DB 271 GTAAGTGAATTCATGAGGAAATCATGGAAGTGAATTAATCTAGTCTTAAAAAGTTAA 330
 |||
 QY 308 GGAGCAGAAAAATTCAGATTCGAAACATCCCTCTCACTGACGTGGAGAGTGTGGAGT 367
 |||
 DB 331 GGAGCAGAAAAATTCAGATTCGAAACATCCCTCTCACTGACGTGGAGAGTGTGGAGT 390
 |||
 QY 368 GACTTTTGGCTCAATATGGGACAGTGAAGATGTGAAACAGTCAACACAGACAGAAA 427
 |||
 DB 391 GACTTTTGGCTCAATATGGGACAGTGAAGATGTGAAACAGTCAACACAGACAGAAA 450
 |||
 QY 428 CGCCGCTTGTCAACGTCACTATGCAACAGAGAAAGAAAGAAATATGCAATGAGAGC 487
 |||
 DB 451 CGCCGCTTGTCAACGTCACTATGCAACAGAGAAAGAAAGAAATATGCAATGAGAGC 510
 |||
 QY 488 TAAAGCGGAGATCAGTTTGAGAACTACTCTTCAAGATTTCTCAATCCGAGTGAAGAG 547

DB 511 TAAAGCGGATCAGTTTGAAGAACTACTCTTCAAGATTTCTCAATCCGAGTGAAGAG 570
 |||
 QY 548 TGAAGTCCCTTCGCCCCCTCAGCGAGCCAGCGTGGGAGCACTCTTCCGGAGAGCAAG 607
 |||
 DB 571 TGAAGTCCCTTCGCCCCCTCAGCGAGCCAGCGTGGGAGCACTCTTCCGGAGAGCAAG 630
 |||
 QY 608 GGCAGGCCCCCTGGGGGCACTTCTCAGGCGCAACAGATGATTTCCGCTGGGATCTCGG 667
 |||
 DB 631 GGCAGGCCCCCTGGGGGCACTTCTCAGGCGCAACAGATGATTTCCGCTGGGATCTCGG 690
 |||
 QY 668 TCCCAACCAATTTGTTGGTGCATCATCGGAAAGGAGGCTCTGACCATTAAGAACATCA 727
 |||
 DB 691 TCCCAACCAATTTGTTGGTGCATCATCGGAAAGGAGGCTCTGACCATTAAGAACATCA 750
 |||
 QY 728 CTAAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAAAGAACTCTGAGCTGACAGAGA 787
 |||
 DB 751 CTAAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAAAGAACTCTGAGCTGACAGAGA 810
 |||
 QY 788 AGCCTGTACCATCATGCCACCCCAAGGGGACTTTGAAAGCATGCCGATGATTCCTG 847
 |||
 DB 811 AGCCTGTACCATCATGCCACCCCAAGGGGACTTTGAAAGCATGCCGATGATTCCTG 870
 |||
 QY 848 AATATCATGAGAAAGAGGAGATGAGACCAACTAGCGGAGAGATTCCTGAAATCT 907
 |||
 DB 871 AATATCATGAGAAAGAGGAGATGAGACCAACTAGCGGAGAGATTCCTGAAATCT 930
 |||
 QY 908 TGGCACAATGCGTTGGTGAAGAGCTGATTTGAAAAAGAGGAGAAATTTGAAGAAA 967
 |||
 DB 931 TGGCACAATGCGTTGGTGAAGAGCTGATTTGAAAAAGAGGAGAAATTTGAAGAAA 990
 |||
 QY 968 TTGAACATGAAACAGGAGCAAGATTAACAATCTCATCTTTGACAGATTGAGCATATACA 1027
 |||
 DB 991 TTGAACATGAAACAGGAGCAAGATTAACAATCTCATCTTTGACAGATTGAGCATATACA 1050
 |||
 QY 1028 ACCCGGAAAGAACATGACTGTGAAGGCGACAGTTGAGGCTGCGCCAGATGAGATAG 1087
 |||
 DB 1051 ACCCGGAAAGAACATGACTGTGAAGGCGACAGTTGAGGCTGCGCCAGATGAGATAG 1110
 |||
 QY 1088 AGATTATGAAGAGCTGCTGAGAGCCTTTAAAAATGATAGCTGTTAAC----- 1140
 |||
 DB 1111 AGATTATGAAGAGCTGCTGAGAGCCTTTAAAAATGATAGCTGTTAACCAAG 1170
 |||
 QY 1141 ----- 1140
 |||
 DB 1171 CCAATGTGATCCAGGGTGAACCTCAGCGCACTTGGCATCTTTCAACAGAGCTGCG 1230
 |||
 QY 1141 ----- 1140
 |||
 DB 1231 TGCATATCCACAGAGAGGCCCCGGAGAGTCCCCCGCTGCCCCCTACACCCCTTCA 1290
 |||
 QY 1141 --ACCACTCCGAGATCTTCTCCAGCTGTACCCCAATCAACAGTTGACCGTCCGC 1198
 |||
 DB 1291 CTACCACTCCGAGATCTTCTCCAGCTGTACCCCAATCAACAGTTGACCGTCCGC 1350
 |||
 QY 1199 ATCATCACTTTATCCAGACAGAGATGTGATTCCTTCAATCCAAACCCAGCTGTG 1258
 |||
 DB 1351 ATCATCACTTTATCCAGACAGAGATGTGATTCCTTCAATCCAAACCCAGCTGTG 1410
 |||
 QY 1259 GGCACATCATCGGAGAAAGAGGGGCAACATCAAAACAGCTGGGAGATTCGCGAGCCT 1318
 |||
 DB 1411 GGCACATCATCGGAGAAAGAGGGGCAACATCAAAACAGCTGGGAGATTCGCGAGCCT 1470
 |||
 QY 1319 CTATCAAGATTGCCCTGGGAGAGGCCAGAGCTCAGCGAAAGATGTCATCATCACCG 1378
 |||
 DB 1471 CTATCAAGATTGCCCTGGGAGAGGCCAGAGCTCAGCGAAAGATGTCATCATCACCG 1530
 |||
 QY 1379 GGCACACCGAGAGCTCGATTCAGAGCCCAAGGAGCGATCTTTGGGAAATCGAAAAGAGAAA 1438
 |||
 DB 1531 GGCACACCGAGAGCTCGATTCAGAGCCCAAGGAGCGATCTTTGGGAAATCGAAAAGAGAAA 1590
 |||
 QY 1439 ACTTCTTAAACCCCAAGAGAGATGAGCTGGAACGCAATATCAAGTGGCCCTCTTCA 1498
 |||

```

Db 1591 ACTCTTTAACCCCAAGAGAGAGTAGAGTGAASGCGATATGAGATGCCCTCTTCCA 1650
QY 1499 CAGCTGCGCGGATGATGGCAAAAGGTGCAAGACCGTGAACGAACTGCAAGACTTAACCA 1558
Db 1551 CAGCTGCGCGGATGATGGCAAAAGGTGCAAGACCGTGAACGAACTGCAAGACTTAACCA 1710
QY 1559 GTGCAAGATGATGCTGCTGCTGCAAAAGGCGATGAATGAGGAATGATGCTCA 1618
Db 1711 GTGCAAGATGATGCTGCTGCTGCAAAAGGCGATGAATGAGGAATGATGCTCA 1770
QY 1619 GAATTAATGAGGATCTCTTTGCTAGCCAGATGCAACGCGCAAGATCAGGAAATTTGAC 1678
Db 1771 GAATTAATGAGGATCTCTTTGCTAGCCAGATGCAACGCGCAAGATCAGGAAATTTGAC 1830
QY 1679 AACAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1738
Db 1831 AACAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1890
QY 1739 GAGGCTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1798
Db 1891 GAGGCTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1950
QY 1799 AATGAGACCAAAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1858
Db 1951 AATGAGACCAAAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2010
QY 1859 AAGTCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1918
Db 2011 AAGTCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2070
QY 1919 GGGCGGGGGAAGGTCAAGCCAGAGTGTGCAAGAACCAAGAGAGAGAGAGAGAGAGAG 1978
Db 2071 GGGCGGGGGAAGGTCAAGCCAGAGTGTGCAAGAACCAAGAGAGAGAGAGAGAGAGAG 2130
QY 1979 GCTTCTGCAAGGCTTCAAGCCATCCATCCATCCATCCATCCATCCATCCATCCATCC 2038
Db 2131 GCTTCTGCAAGGCTTCAAGCCATCCATCCATCCATCCATCCATCCATCCATCCATCC 2190
QY 2039 GAGCGTATCCCTTTTGTGGAATTAATGAGTGAACGCTGTTCAAGCGCAAGAAATGC 2098
Db 2191 GAGCGTATCCCTTTTGTGGAATTAATGAGTGAACGCTGTTCAAGCGCAAGAAATGC 2250
QY 2099 AACACCTTTTTCGTGAGCAATGCTCTGTAATGATGATGATGATGATGATGATGATGATG 2158
Db 2251 AACACCTTTTTCGTGAGCAATGCTCTGTAATGATGATGATGATGATGATGATGATGATG 2310
QY 2159 TGTTAAGATATGAGCTGTGAGTGAACAGAGGTGCTGACGAGTATATATTTAGAA 2218
Db 2311 TGTTAAGATATGAGCTGTGAGTGAACAGAGGTGCTGACGAGTATATATTTAGAA 2370
QY 2219 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2278
Db 2371 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2430
QY 2279 TTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2338
Db 2431 TTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2490
QY 2431 TTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2490
Db 2491 GGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2550
QY 2339 GGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2398
Db 2551 GGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2610
QY 2399 GAGAGACACTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2458
Db 2551 GAGAGACACTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2610
QY 2459 CTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2518
Db 2611 CTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2670
QY 2519 GCTTCAAGGATTTTAACTGATTTTGTGATGATGATGATGATGATGATGATGATGATG 2578
Db 2671 GCTTCAAGGATTTTAACTGATTTTGTGATGATGATGATGATGATGATGATGATGATG 2730

```

```

QY 2579 TTTATCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2638
Db 2731 TTTATCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2790
QY 2639 ATCCCTCATCTCAATCCCGATATGAGCA--CCCCCCCCCCCCCAAGGCAAAAGAGTCT 2696
Db 2791 ATCCCTCATCTCAATCCCGATATGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2850
QY 2697 CTGAGTATCATATCAACAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2756
Db 2851 CTGAGTATCATATCAACAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2910
QY 2757 ACTTGTTACTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2816
Db 2911 ACTTGTTACTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2970
QY 2817 AACAGAAACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 2876
Db 2971 AACAGAAACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 3030
QY 2877 ATGATTTTGTCTTTTGTGATTTGATGATGATGATGATGATGATGATGATGATGATGATG 2936
Db 3031 ATGATTTTGTCTTTTGTGATTTGATGATGATGATGATGATGATGATGATGATGATGATG 3090
QY 2937 TTTCTTT--AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2995
Db 3091 TTTCTTTAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3150
QY 2996 GGTAAATTCACAAATTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3055
Db 3151 GGTAAATTCACAAATTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3210
QY 3056 TTTTACTCAGACCTGCTCTGTTTCCCTTGAAGATTTGTAAAGCTGATGATG 3115
Db 3211 TTTTACTCAGACCTGCTCTGTTTCCCTTGAAGATTTGTAAAGCTGATGATGATG 3270
QY 3116 AGCATTTTATATTTTATATATATATATATATATATATATATATATATATATATATAT 3175
Db 3271 AGCATTTTATATTTTATATATATATATATATATATATATATATATATATATATATAT 3330
QY 3176 GCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3235
Db 3331 GCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3390
QY 3236 AAGAACCAATATGAGCTTCTTTTGAACAAACCTTGAATTTATTT 3283
Db 3391 AAGAACCAATATGAGCTTCTTTTGAACAAACCTTGAATTTATTT 3438

RESULT 7
AAS26153
ID AAS26153 standard; cDNA; 2290 BP.
AC AAS26153;
DT 07-NOV-2001 (first entry)
XX
DE Human cDNA encoding a novel secreted protein, Seg ID 332.
KW Human; immunosuppressive; antiarthritic; ss; antirheumatic; cytostatic;
KW cardiac; vasotropic; cerebroprotective; neurotropic; neuroprotective;
KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN W0200155322-A2.

```


DR WPI: 2001-488783/53.
DR P-PSDB; AAIU6166.

XX New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.

PS Claim 1; SEQ ID NO 332; 980bp; English.

XX The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Antibodies to the proteins can also be used in
CC alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. Rheumatoid arthritis, hyperproliferative disorders e.g. cardiac
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC and many other disorders listed in the specification. The polypeptides
CC can also be used to aid wound healing and epithelial cell proliferation,
CC to prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence encodes a novel secreted protein of the invention. Note: The
CC sequence data for this patent did not form part of the printed

Query Match 60.6%; Score 1989.8; DB 4; Length 2290;
Best Local Similarity 99.4%; Pred. No. 0;

Matches 2028; Conservative 0; Mismatches 8; Indels 4; Gaps 3;

QY	1247	CCCAAGCTGTGGGCGCATCATCGGGAAGAGGGGGGACACATCAACAGCTGGCGAGAT	1306
DB	25	CCCAAGCTGTGGGCGCATCATCGGGAAGAGGGGGGACACATCAACAGCTGGCGAGAT	84
QY	1307	TCGCGGAGGCTTATCAAGATTGCGCTGCGAAGGCCCAAGCTGACGGAAAGATGG	1366
DB	85	TCGCGGAGGCTTATCAAGATTGCGCTGCGAAGGCCCAAGCTGACGGAAAGATGG	144
QY	1367	TCATCATCAACGGGCGACCGGAGGCCGATTCAGGCGGAGCGATTTGGGAAAC	1426
DB	145	TCATCATCAACGGGCGACCGGAGGCCGATTCAGGCGGAGCGATTTGGGAAAC	204
QY	1427	TGAAGAAGAGAACTTCTTACCCCAAGAGAGTGAAGTGAAGCGCATATCAGAG	1486
DB	205	TGAAGAAGAGAACTTCTTACCCCAAGAGAGTGAAGTGAAGCGCATATCAGAG	264
QY	1487	TGCGCTCTTCCACAGCTGCGCGGATGTTGGCAAAAGGTGCGAAGCCGTGAAGACTGC	1546
DB	265	TGCGCTCTTCCACAGCTGCGCGGATGTTGGCAAAAGGTGCGAAGCCGTGAAGACTGC	324
QY	1547	AGAACTTAACAGTGAAGTCACTGCTGCTGTCGCAAAAGCGGAGTGAAGAAATGAG	1606
DB	325	AGAACTTAACAGTGAAGTCACTGCTGCTGTCGCAAAAGCGGAGTGAAGAAATGAG	384
QY	1607	AAGTGAATGTCAGAAATATGCGGCACTTTTGTGTAACAGACTGACAGCGCAAGTCA	1666
DB	385	AAGTGAATGTCAGAAATATGCGGCACTTTTGTGTAACAGACTGACAGCGCAAGTCA	444
QY	1667	GGGAAATTTGTAACAAGGTGAAGAGCAAGAAATACCTTCAGGGAGTGCCTTCAC	1726
DB	445	GGGAAATTTGTAACAAGGTGAAGAGCAAGAAATACCTTCAGGGAGTGCCTTCAC	504
QY	1727	AGGCGACGAAGTGGCTCCCAAGGCAACAGCAAAACAAGATGAATGAAGCCCTTCC	1786
DB	505	AGGCGACGAAGTGGCTCCCAAGGCAACAGCAAAACAAGATGAATGAAGCCCTTCC	564

QY	1787	AACACCTGACAGATGAGACCAAAAGCGACCGACCGATCGGAGGCAAAACAAAGACAT	1846
DB	565	AACACCTGACAGATGAGACCAAAAGCGACCGACCGATCGGAGGCAAAACAAAGACAT	624
QY	1847	CTGAGAAATGAAGATCTCTCGGAGGCGGCGGAGGAGTCTCGCGAGGCGCTGGAACCCCA	1906
DB	625	CTGAGAAATGAAGATCTCTCGGAGGCGGCGGAGGAGTCTCGCGAGGCGCTGGAACCCCA	684
QY	1907	GGGCGCGAGAGGCGGCGGAGGAGTCAAGCGAGTTTCCAGAAACCAACCGAGCCCGCTTC	1966
DB	685	GGGCGCGAGAGGCGGCGGAGGAGTCAAGCGAGTTTCCAGAAACCAACCGAGCCCGCTTC	744
QY	1967	CGCGCCCGGAGGCGGCGGAGGAGTCAAGCGAGTTTCCAGAAACCAACCGAGCCCGCTTC	2026
DB	745	CGCGCCCGGAGGCGGCGGAGGAGTCAAGCGAGTTTCCAGAAACCAACCGAGCCCGCTTC	804
QY	2027	CTGAATCTCCACAGACGCTATCCCTTTTAACTGAACATACATAGGTAGCGTTCAAGC	2086
DB	805	CTGAATCTCCACAGACGCTATCCCTTTTAACTGAACATACATAGGTAGCGTTCAAGC	864
QY	2087	CAAGCAAAATGACACACCTTTTCTGTGCAAAATGCTCTGTATACATGTGTATATAT	2146
DB	865	CAAGCAAAATGACACACCTTTTCTGTGCAAAATGCTCTGTATACATGTGTATATAT	924
QY	2147	AGAAAGGAGAAATGTTAAGATATGTGGCTGTGGTTTACACAGGCTGCGAGCGTTAA	2206
DB	925	AGAAAGGAGAAATGTTAAGATATGTGGCTGTGGTTTACACAGGCTGCGAGCGTTAA	984
QY	2207	TATATTTTGAAT	2266
DB	985	TATATTTTGAAT	1044
QY	2267	ATTTTCTTTCTTTTAAAGAGAAACAGGCTTTTCTGAGCTTTTAAAGATTAAGCTTT	2326
DB	1045	ATTTTCTTTCTTTTAAAGAGAAACAGGCTTTTCTGAGCTTTTAAAGATTAAGCTTT	1104
QY	2327	GGGAGTCTCAAGGTTGAGAGAGAGCTTTGAGGCGACCGGCAAAATTCACCCAGAG	2386
DB	1105	GGGAGTCTCAAGGTTGAGAGAGAGCTTTGAGGCGACCGGCAAAATTCACCCAGAG	1164
QY	2387	GAAATCTCGTGGAGAGAGACCTGACGCGAGTTGATGATCACTGTGTATATCAAGAG	2446
DB	1165	GAAATCTCGTGGAGAGAGACCTGACGCGAGTTGATGATCACTGTGTATATCAAGAG	1224
QY	2447	GATATACGCTCTCTTGAAGAGAGAACTGTGTCACTCTCATGCTGTAGTCAATAC	2506
DB	1225	GATATACGCTCTCTTGAAGAGAGAACTGTGTCACTCTCATGCTGTAGTCAATAC	1284
QY	2507	CAATTTCTCTTCTTCAAGGTTTAACTGTTTGTGATATCTGTATATATATCTC	2566
DB	1285	CAATTTCTCTTCTTCAAGGTTTAACTGTTTGTGATATCTGTATATATATCTC	1344
QY	2567	TGTCTCTCTCTTCTTCAAGGTTTAACTGTTTGTGATATCTGTATATATATCTC	2626
DB	1345	TGTCTCTCTCTTCTTCAAGGTTTAACTGTTTGTGATATCTGTATATATATCTC	1403
QY	2627	TGAATTTCTCTTCAAGGTTTAACTGTTTGTGATATCTGTATATATATCTC	2684
DB	1404	TGAATTTCTCTTCAAGGTTTAACTGTTTGTGATATCTGTATATATATCTC	1463
QY	2685	CAAGAGTGTCTGAGTATCAATCAACAAAGAGAAACAAAGCGGAAACACACAAACA	2744
DB	1464	CAAGAGTGTCTGAGTATCAATCAACAAAGAGAAACAAAGCGGAAACACACAAACA	1523
QY	2745	GCTCAACTTACACTGGTACTCAAAAGAGAGAGTCAATGGTACTTCTAGAGCTT	2804
DB	1524	GCTCAACTTACACTGGTACTCAAAAGAGAGAGTCAATGGTACTTCTAGAGCTT	1583
QY	2805	TTGGAAGAGAGAAACAGAAACCAACCAACCAATCAACCAACCAAGAAATTC	2864
DB	1584	TTGGAAGAGAGAAACAGAAACCAACCAACCAATCAACCAACCAAGAAATTC	1643

QY 2865 CACAAATGAAAGATGATATTTGCTTTTGGATTTGGTATTAAGCCATCAATATTCAG 2924
DB 1644 CACAAATGAAAGATGATATTTGCTTTTGGATTTGGTATTAAGCCATCAATATTCAG 1703
QY 2925 CAAATGATTCCTTTCTTTT-AAAAAATAATGAGGAAAGTAAATTTACCAAGT 2983
DB 1704 CAAATGATTCCTTTCTTTTAAAAAATAATGAGGAAAGTAAATTTACCAAGT 1763
QY 2984 TGTGTGCCCCAGGGCGCTTAAATTCAGAGATTTTAAACGAGAAAAACACAGAAAGAC 3043
DB 1764 TGTGTGCCCCAGGGCGCTTAAATTCAGAGATTTTAAACGAGAAAAACACAGAAAGAC 1823
QY 3044 TACCTCAGAGTGTTTTACCTCAGACCTTGCTTGTGTTCCCTTAGAGATTTGTAA 3103
DB 1824 TACCTCAGAGTGTTTTACCTCAGACCTTGCTTGTGTTCCCTTAGAGATTTGTAA 1883
QY 3104 GCTGATGATTTGAGACATTTTATTTTATTTTAAATTAATAATGATTTGAAAAATAAGA 3163
DB 1884 GCTGATGATTTGAGACATTTTATTTTATTTTAAATTAATAATGATTTGAAAAATAAGA 1943
QY 3164 TATCAACTGCGCAGCTGAGAGAGTGAAGTCCAAAGTGTGAACAGCTGTTCTGAATTGT 3223
DB 1944 TATCAACTGCGCAGCTGAGAGAGTGAAGTCCAAAGTGTGAACAGCTGTTCTGAATTGT 2003
QY 3224 CTTCGGCTAGCCAGAAACCNATATGSCCTTCTTTTGGACAAACCTTGAAAAATGTTATTT 3283
DB 2004 CTTCGGCTAGCCAGAAACCNATATGSCCTTCTTTTGGACAAACCTTGAAAAATGTTATTT 2063

RESULT 8
ABX73494
ID ABX73494 standard; DNA; 2290 BP.
AC ABX73494;
XX
DT 18-MAR-2003 (first entry)
XX
DE Human novel polynucleotide #322.
XX
KW Human; gene; ds; neural disorder; immune system disorder; renal disorder;
KW muscular disorder; respiratory disease; reproductive disorder;
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW hyperproliferative disorder; inflammatory disease; allergic reaction;
KW blood related disorder; cancer; immunosuppressive; antinflammatory;
KW cardiovascular; nephrotoxic; cytostatic; antiallergic; chromolytic;
KW haemostatic; antiatherosclerotic.
OS Homo sapiens.
XX
XX
PN US2002132753-A1.
XX
PD 19-SEP-2002.
XX
PF 17-JAN-2001; 2001US-00764864.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.

PR 22-AUG-2000; 2000US-0226866P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249239P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
XX
XX (ROSE/) ROSEN C A.
PA (RUBEN/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
XX WPI; 2003-14744/14.
DR P-PSDB; ABUS5234.
XX
PT New polypeptides and nucleic acids, useful in gene therapy for treating,
PT inhibiting or preventing e.g. neural, immune system, muscular,
PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT renal disorders.
XX
PS Claim 1; SEQ ID NO 332; 402pp; English.
XX
XX The invention relates to human novel polypeptides and their associated
CC polynucleotides. The polypeptides and polynucleotides are useful in gene
CC therapy for treating, inhibiting or preventing neural disorders, immune
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC leukemia), inflammatory diseases (e.g. septic shock, bursitis and
CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent
CC human novel polynucleotides of the invention
XX
SQ Sequence 2290 BP; 707 A; 561 C; 491 G; 531 T; 0 U; 0 Other;
Query Match 60.6%; Score 1989.8; DB 7; Length 2290;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 2028; Conservative 0; Mismatches 8; Indels 4; Gaps 3;
QY 1247 CCAGGCTGTGGGCGCATCTGCGAAGAGAGGGGCGCACATCAACAGCTGGCAGAT 1306
|||||

Db 25 CCGAAGGCTCCGGGCCCATCATCGGGAAGAGGGGCAACATCAACAGCTGGCAGAT 84
QY 1307 TCGCCGAGGCTCTATCAAGATTGCCCTCGGGAAGGCCAGACGTCAAGCCGAAAGATGG 1366
Db 85 TCGCCGAGGCTCTATCAAGATTGCCCTCGGGAAGGGCCAGACGTCAAGCCGAAAGATGG 144
QY 1367 TCATCATCAACGGGGCCACCGGAAGCCAGTTCAAGGCCAGGGAAGGATCTTTGGGAAAC 1426
Db 145 TCATCATCAACGGGGCCACCGGAAGCCAGTTCAAGGCCAGGGAAGGATCTTTGGGAAAC 204
QY 1427 TGAAGAGGAAAACTCTTTAAACCCAAAAGAAAGTGAAGCTGGAAGCGCATATCAAG 1486
Db 205 TGAAGAGGAAAACTCTTTAAACCCAAAAGAAAGTGAAGCTGGAAGCGCATATCAAG 264
QY 1487 TGCCTCTTCCACAGCTGGCCGGGTGATTGGCAAAAGTGGCAAGCCGGAAAGAACTGC 1546
Db 265 TGCCTCTTCCACAGCTGGCCGGGTGATTGGCAAAAGTGGCAAGCCGGAAAGAACTGC 324
QY 1547 AGAAGTTAAACGAGTGAAGTCAATCGTGCCTCGTGAACCAAGCCAGATGAAGAAATGAG 1606
Db 325 AGAAGTTAAACGAGTGAAGTCAATCGTGCCTCGTGAACCAAGCCAGATGAAGAAATGAG 384
QY 1607 AAGTATCGTCAGAAATTAATCGGCACTTCTTTGTAACCACTGCAACGCCAAGATTA 1666
Db 385 AAGTATCGTCAGAAATTAATCGGCACTTCTTTGTAACCACTGCAACGCCAAGATTA 444
QY 1667 GGGAAATTGTAAACAGGTGAAGCAGCAGAGCAGAAATACCTCAAGGGAGTGCCTCAC 1726
Db 445 GGGAAATTGTAAACAGGTGAAGCAGCAGAGCAGAAATACCTCAAGGGAGTGCCTCAC 504
QY 1727 ACGCGACGAAGTGAAGGCTCCACAGGCAACAGCAAAACACGATGATGATGAGCCCTTC 1786
Db 505 ACGCGACGAAGTGAAGGCTCCACAGGCAACAGCAAAACACGATGATGATGAGCCCTTC 564
QY 1787 AACACCTGAACAGATGAACAAAACGACCCAGCCAGATCGGAGCAACAAAGACCT 1846
Db 565 AACACCTGAACAGATGAACAAAACGACCCAGCCAGATCGGAGCAACAAAGACCT 624
QY 1847 CTGAGGAATGAAGTCTCGGAGGGGGCAGGGAATCTGCGGAGGCCCTGAGAACCCCA 1906
Db 625 CTGAGGAATGAAGTCTCGGAGGGGGCAGGGAATCTGCGGAGGCCCTGAGAACCCCA 684
QY 1907 GGGGCGCAGAGGGGGCGGGGAAGTCAAGCAGGTTTGCAGAACCAACGAGCCCGCTC 1966
Db 685 GGGGCGCAGAGGGGGCGGGGAAGTCAAGCAGGTTTGCAGAACCAACGAGCCCGCTC 744
QY 1967 CCGCCCCCAGGGCTTCTGAGGCTTCAAGCATCACTTCAACATCACTCGGATCTCTC 2026
Db 745 CCGCCCCCAGGGCTTCTGAGGCTTCAAGCATCACTTCAACATCACTCGGATCTCTC 804
QY 2027 CTGAACTCCACAGACGTATCCCTTTTAACTTGAATTAACATAGTGAAGCTGTTCAAGC 2086
Db 805 CTGAACTCCACAGACGTATCCCTTTTAACTTGAATTAACATAGTGAAGCTGTTCAAGC 864
QY 2087 CAAGCAAAATGACACACCTTTTCTGTGCAAAATCGTCTCTGTACATGTGTGATATTT 2146
Db 865 CAAGCAAAATGACACACCTTTTCTGTGCAAAATCGTCTCTGTACATGTGTGATATTT 924
QY 2147 AGAAGGGAAGATGTTAAGATATGTGCGCTGTGGGTTAACAAGGTGCTGCAAGGTTA 2206
Db 925 AGAAGGGAAGATGTTAAGATATGTGCGCTGTGGGTTAACAAGGTGCTGCAAGGTTA 984
QY 2207 TATATTTTAAATTAATATATCAATTAACATCACTCAATTTTAACTTAACTTATTA 2266
Db 985 TATATTTTAAATTAATATATCAATTAACATCACTCAATTTTAACTTAACTTATTA 1044
QY 2267 AATTTTTTTCTTTTAAAGAAAGCAGCTTTCTAGACTTTAAAGAAATAAGTCTTT 2326
Db 1045 AATTTTTTTCTTTTAAAGAAAGCAGCTTTCTAGACTTTAAAGAAATAAGTCTTT 1104
QY 2327 GGGAGGTCTCAACGGTGTAGAGAGAGCTTTGAGGCCACCGCACAATAATTCACCAAGG 2386
Db 1105 GGGAGGTCTCAACGGTGTAGAGAGAGCTTTGAGGCCACCGCACAATAATTCACCAAGG 1164

QY 2387 GAAATCTGTGGAAGGACACTCAAGCAGTCTGAGTCACTGTGTATGTCAACAGAG 2446
Db 1165 GAAATCTGTGGAAGGACACTCAAGCAGTCTGAGTCACTGTGTATGTCAACAGAG 1224
QY 2447 GATATCGGTCTCTTGAAGAGAAACTGTATCACTCCCAAGCCGTAGTCAATAC 2506
Db 1225 GATATCGGTCTCTTGAAGAGAAACTGTATCACTCCCAAGCCGTAGTCAATAC 1284
QY 2507 CCATTTCTCTTGTCTCAACAGTCTTAACTGATTTTGTGCACTGCTATATTAATCTC 2566
Db 1285 CCATTTCTCTTGTCTCAACAGTCTTAACTGATTTTGTGCACTGCTATATTAATCTC 1344
QY 2567 TGTCTCTCTGTATATCTCTCCCTCCCTCCCTCCCTCTTCTCCATCTTCAATCTT 2626
Db 1345 TGTCTCTCTGTATATCTCTCCCTCCCTCCCTCCCTCTTCTCCATCTTCAATCTT 1403
QY 2627 TTGAATTTCTCATCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 2684
Db 1404 TTGAATTTCTCATCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1463
QY 2685 CAAAGCAGTCTGTAGTATCATATCAACAAAGGAAAGGAAAGGAAAGGAAAGGAAAG 2744
Db 1464 CAAAGCAGTCTGTAGTATCATATCAACAAAGGAAAGGAAAGGAAAGGAAAGGAAAG 1523
QY 2745 GCCTCACTTACACTTGTACTCAAAAGACAAAGATCAATGTACTTGTCTTACGCTT 2804
Db 1524 GCCTCACTTACACTTGTACTCAAAAGACAAAGATCAATGTACTTGTCTTACGCTT 1583
QY 2805 TTGGAAGAGGAAACAGAAAGCCACCAACCAACCAATCAACCAACCAAGAAATATC 2864
Db 1584 TTGGAAGAGGAAACAGAAAGCCACCAACCAACCAATCAACCAACCAAGAAATATC 1643
QY 2865 CACAATGAAGAATATTTGTCTTTTGAATTTGTGTATAGCAATCAATATTCAG 2924
Db 1644 CACAATGAAGAATATTTGTCTTTTGAATTTGTGTATAGCAATCAATATTCAG 1703
QY 2925 CAAATGATTCCTTTCTTT-AAAAAATAATGTGAGGAAATGAAATTTTCCAAGGT 2983
Db 1704 CAAATGATTCCTTTCTTTAAAAAATAATGTGAGGAAATGAAATTTTCCAAGGT 1763
QY 2984 TGTGGCCCAAGGCGTTAATTCAGATTTTAAAGGAAACACACAGAAAGAGC 3043
Db 1764 TGTGGCCCAAGGCGTTAATTCAGATTTTAAAGGAAACACACAGAAAGAGC 1823
QY 3044 TACCTAGAGTGTATTTTACCTAGACACTGCTGTGTGTTTCTTAAAGATTTGTAA 3103
Db 1824 TACCTAGAGTGTATTTTACCTAGACACTGCTGTGTGTTTCTTAAAGATTTGTAA 1883
QY 3104 GCTGATAGTTGAGCAATTTTATTTTAAATTAATTAATGAGTGGAAAAAATAAGA 3163
Db 1884 GCTGATAGTTGAGCAATTTTATTTTAAATTAATTAATGAGTGGAAAAAATAAGA 1943
QY 3164 TATCACTGCGACCCGGAAGAGTGAAGTCAAGTGTCAACAGCTGTTCAATTTGT 3223
Db 1944 TATCACTGCGACCCGGAAGAGTGAAGTCAAGTGTCAACAGCTGTTCAATTTGT 2003
QY 3224 CTTCGCCGTAAGCCAAAGACNAATATGAGCTTTTGTGACAAACCTTGAATTTATTT 3283
Db 2004 CTTCGCCGTAAGCCAAAGACNAATATGAGCTTTTGTGACAAACCTTGAATTTATTT 2063

RESULT 9
ID AAS70981 standard; cDNA; 2010 BP.
XX AAS70981;
AC 13-FEB-2002 (first entry)
DT 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #6785.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW

XX	food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US008631.
XX	
PR	31-MAR-2000; 2000US-00540217.
PR	23-AUG-2000; 2000US-00649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Drmnac RT, Liu C, Tang YT;
XX	
DR	WPI; 2001-639362/73.
DR	P-PSDB, ABG06794.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity.
XX	
SS	Claim 1; SEQ ID NO 6785; 103bp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantifying a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. Ab664197-Ab834564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences

Sequence 2010 BP; 565 A; 552 C; 532 G; 361 T; 0 U; 0 Other;

Query Match	52.1%	Score 171.8	DB 5	length 2010
Best Local Similarity	92.9%	Pred. No. 0		
Matches 1858; Conservative	0	Mismatches 12	Indels 129	Gaps 1

QY	3	GCCGGGGGACCGCGGGGCTCTCGGGGAGACCGATGATGAACAGACTTAACTCGGGA	94
Db	12	GCCGTATGGAGCCCGGCTCATGGGGGAGAGACGATATGAAACAACTTTAATCGGG	71
QY	95	ACCTGAGCCCGCCCTCATCCGCGACACTTTTGGGGACAGGAAGCTGC	154
Db	72	ACCTAGACCCCGCCCTCATCCGCGACACTCTTTGGGGACAGGAAGCTGC	131
QY	155	CCCTGGCGGACAGGTCTCTGTGAAGTCCGCGCTACGCGCTTGATGATACCCGACGACA	214
Db	132	CCCTGGCGGACAGGTCTCTGTGAAGTCCGCGCTTGATGATACCCGACGACA	191
QY	215	ACCGGGCCATCCGGCGCATCGAGACCCCTTCGGGTAAAGTGAATTGCATGGGAAAATCA	274
Db	192	ACCGGGCCATCCGGCGCATCGAGACCCCTTCGGGTAAAGTGAATTGCATGGGAAAATCA	251
QY	275	TGGAGTTGATTACTCACTCTCTAAANAAGCTTAAGAGACGAGAAATTCAGATTCGAAACA	334

Db	252	TGGAAAGTTGATTAACTGACGCTCTGTAAAAAGCTAAGGACGAGAAAAATTCAGATTTCGAAAA	311
QY	335	TCCCTCTTCACCTGAGAGTGGAGAGTGTGGATGACCTTTGGCTCAATATGAGGACAGTGG	394
Db	312	TCCCTCTTCACCTGAGAGTGGAGAGTGTGGATGACCTTTGGCTCAATATGAGGACAGTGG	371
QY	395	AGAAATGTGGAAACAAGTCAACACAGACACAGAAACCGCGTGTCAACGTCAACATATGACA	454
Db	372	AGAAATGTGGAAACAAGTCAACACAGACACAGAAACCGCGTGTCAACGTCAACATATGACA	431
QY	455	CAAGGAGAAAGCAAAAAATAGCATATGAGAAAGCTAAGCGGCTATCACTTTGAGAACTACT	514
Db	432	CAAGGAGAAAGCAAAAAATAGCATATGAGAAAGCTAAGCGGCTATCACTTTGAGAACTACT	491
QY	515	CCTTAAAGATTCTTACATCCCGAGTGAAGAGGTAGGCTCCCTTTGGCCCTTCAGCGAG	574
Db	492	CCTTAAAGATTCTTACATCCCGAGTGAAGAGGTAGGCTCCCTTTGGCCCTTCAGCGAG	551
QY	575	CCGACGCTGGGGACCACTCTTCCGCGGAGCAAGGCAAGCCCTGGGGGACCTTCTCAGG	634
Db	552	CCGACGCTGGGGACCACTCTTCCGCGGAGCAAGGCAAGCCCTGGGGGACCTTCTCAGG	611
QY	635	CCAGACAATATGATTTCCCGCTGGGGATCTGTGTCTCCCATCCCAAGTTTGTGTGTGCATCA	694
Db	612	CCAGACAATATGATTTCCCGCTGGGGATCTGTGTCTCCCATCCCAAGTTTGTGTGTGCATCA	671
QY	695	TCGGAAAGAGAGGCTTGAACATATAAGAACATCACTAAGCAGACCCAGTCCGGGTATGATA	754
Db	672	TCGGAAAGAGAGGCTTGAACATATAAGAACATCACTAAGCAGACCCAGTCCGGGTATGATA	731
QY	755	TCATATGAAAAAGAACTCTGAGCTGCAGAAAGCCTGTCAACCATCATCCACCCAG	814
Db	732	TCATATGAAAAAGAACTCTGAGCTGCAGAAAGCCTGTCAACCATCATCCACCCAG	791
QY	815	AGGGGACTTTCGAAGCATGCGCGATGATCTTTGAATCATGCAAGAGGCGAGATGAGA	874
Db	792	AGGGGACTTTCGAAGCATGCGCGATGATCTTTGAATCATGCAAGAGGCGAGATGAGA	851
QY	875	CCAAACTAGCCGAGAGAGATCTCTGAAATCTTGGACACAAATGGCTTGGTTGGAAGAC	934
Db	852	CCAAACTAGCCGAGAGAGATCTCTGAAATCTTGGACACAAATGGCTTGGTTGGAAGAC	911
QY	935	TGATTGAAAAAGAGGAGAAATTTGAAAGAAATTTGAACATGAAACAGGGGACCAAGATA	994
Db	912	TGATTGAAAAAGAGGAGAAATTTGAAAGAAATTTGAACATGAAACAGGGGACCAAGATA	971
QY	995	CAATCTCATCTTTGCAGAGATTTGACATATACAAACCGGAAGAGAACATCATCTGTGAAG	1054
Db	972	CAATCTCATCTTTGCAGAGATTTGACATATACAAACCGGAAGAGAACATCATCTGTGAAG	1031
QY	1055	GCAACAGTTGAGGCGCTGTGCGACAGTGCAGATAGAGATTAGAGAAAGCTCGTAGGCGCT	1114
Db	1032	GCAACAGTTGAGGCGCTGTGCGACAGTGCAGATAGAGATTAGAGAAAGCTCGTAGGCGCT	1091
QY	1115	TTGAAAAATGATATAGCTGGCTGTTAAC-----	1140
Db	1092	TTGAAAAATGATATAGCTGGCTGTTAAC-----	1151
QY	1141	-----	1140
Db	1152	GCGCACTTGGCACTTTTTCACAGAGACTGTCCGTGTATCTCACAGCAGGCGCCGAG	1211
QY	1141	-----ACCACCTCCGATATCTTCCAGCC	1165
Db	1212	GAGTCCCCCGCTGCCCTTACACACCCCTTCACTACCACTCCGATATCTTCTCCAGCC	1271
QY	1166	TGTACCCCAATCAACAGTTTGGCGCGTTCGCGATATCATCTTATACAGACAGAGAGA	1225
Db	1272	TGTACCCCAATCAACAGTTTGGCGCGTTCGCGATATCATCTTATACAGACAGAGAGA	1331
QY	1226	TTTGTGAATCTCTTCATCCCAACCAAGCTGTGGGCGCATCATCGGAAAGAAAGGGGCAAC	1285
Db	1332	TTTGTGAATCTCTTCATCCCAACCAAGCTGTGGGCGCATCATCGGAAAGAAAGGGGCAAC	1391

```

QY 1286 ACATCAACAGCTGGGAGATTCGCCGAGCTCTATCAAGATTGCCCTCGGAAAGCC 1345
Db 1392 ACATCAACAGCTGGGAGATTCGCCGAGCTCTATCAAGATTGCCCTCGGAAAGCC 1451
QY 1346 CAGAGCTGAGGAAAGATGGTCAATCAACGGGGCCAGCCGAAAGCCAGTTCAAGGCC 1405
Db 1452 CAGAGCTGAGGAAAGATGGTCAATCAACGGGGCCAGCCGAAAGCCAGTTCAAGGCC 1511
QY 1406 AGGACGAGATCTTTGGGAACTGAAAGAGAAACCTTTTAACTCCCAAGAGAGTGA 1465
Db 1512 AGGACGAGATCTTTGGGAACTGAAAGAGAAACCTTTTAACTCCCAAGAGAGTGA 1571
QY 1466 AGCTGGAAGCGCATATCAAGTGCCTCTTCCACAGCTGGCGCGGTGATTGGCAAGGTG 1525
Db 1572 AGCTGGAAGCGCATATCAAGTGCCTCTTCCACAGCTGGCGCGGTGATTGGCAAGGTG 1631
QY 1526 GCAAGACGCTGAAGCACTGCAAGACTTAACCAAGTCAAGTCAATCGTGCCTCGTGACC 1585
Db 1632 GCAAGACGCTGAAGCACTGCAAGACTTAACCAAGTCAAGTCAATCGTGCCTCGTGACC 1691
QY 1586 AAACCCAGATGAAATGAGAAATGATCGTCAAGATTATCGGCACTTTCTTGCTAGCC 1645
Db 1692 AAACCCAGATGAAATGAGAAATGATCGTCAAGATTATCGGCACTTTCTTGCTAGCC 1751
QY 1646 AGACTGCAACAGCGCAAGATCAGGGAAATTTGTCACACAGTGAAGCAGCAGCAGAAAT 1705
Db 1752 AGACTGCAACAGCGCAAGATCAGGGAAATTTGTCACACAGTGAAGCAGCAGCAGAAAT 1811
QY 1706 ACCCTCAGGAGTGCCTCAGACGCGCAGCAAGTGAAGTCCCAAGCAGCAGCAAAACA 1765
Db 1812 ACCCTCAGGAGTGCCTCAGACGCGCAGCAAGTGAAGTCCCAAGCAGCAGCAAAACA 1871
QY 1766 ACGATGATGTAGCCCTTCCAAACACTTGAAGATGAGACCAACGCGCAGCCAGAT 1825
Db 1872 ACGATGATGTAGCCCTTCCAAACACTTGAAGATGAGACCAACGCGCAGCCAGAT 1931
QY 1826 CCGGAGCAAAACCAAGAACATCTGAGAGAAATGAGAACTCGCGGAGGCGCAGGAGCTCT 1885
Db 1932 CCGGAGCAAAACCAAGAACATCTGAGAGAAATGAGAACTCGCGGAGGCGCAGGAGCTCT 1991
QY 1886 GCCGAGGCCCTGAGAACCC 1904
Db 1992 GCCGAGGCCCTGAGAACCC 2010

RESULT 10
ID ACA90176 standard; cDNA; 1707 BP.
XX ACA90176;
AC ACA90176;
XX
XX 10-JUL-2003 (first entry)
DT
XX
DE cDNA encoding novel human protein NOV14a.
KW Human; cytosolic; DAPK3-Agonist; DAPK3-Antagonist; cancer; NOV; gene;
XX ss.
XX Homo sapiens.
OS
XX
XX W02003031571-A2.
XX
XX 17-APR-2003.
XX
XX 02-OCT-2002; 2002WO-US031357.
XX
XX 05-OCT-2001; 2001US-0327454P.
XX 09-OCT-2001; 2001US-0327917P.
XX 09-OCT-2001; 2001US-0328029P.
XX 09-OCT-2001; 2001US-0328056P.
XX 12-OCT-2001; 2001US-032849P.
XX 15-OCT-2001; 2001US-0329414P.

```

```

PR 17-OCT-2001; 2001US-0330142P.
PR 22-OCT-2001; 2001US-0341058P.
PR 24-OCT-2001; 2001US-0343629P.
PR 29-OCT-2001; 2001US-0349575P.
PR 01-NOV-2001; 2001US-0346357P.
PR 25-JUN-2002; 2002US-0391342P.
PR 01-OCT-2002; 2002US-00262445.
XX
XX (CUBA-) CUBAGEN CORP.
XX
XX Alsedbrook JP, Burgess CE, Catterton E, Chant JS, Chaudhuri A;
PI Edinger SR, Gerlach VL, Giot L, Gorman L, Guo X, Kekuda R;
PI Mezes PS, Millet I, Ooi CE, Patlurajan M, Rieger DK, Szytek KA,
PI Taupier RJ, Zernusen BD, Zhong H, Zhong M;
XX
XX WPI: 2003-381704/36.
DR P-PsDB; AB089799.
PT
PI New DAPK3 polypeptide, useful for preparing a composition for treating or
XX preventing e.g., cancer.
PS Claim 20; Page 129; 253pp; English.
XX
XX The invention describes an isolated polypeptide comprising any of 33 90-
CC 1273 amino acid sequences (I) given in the specification or its mature
CC form, a sequence that is at least 95 % identical to (I), or a sequence
CC comprising one or more conservative substitutions in the amino acid
CC sequence of (I). The polypeptide is useful for preparing a composition
CC for treating or preventing e.g. cancer. This sequence encodes a novel
CC human NOV protein
XX
XX Sequence 1707 BP; 492 A; 443 C; 442 G; 330 T; 0 U; 0 Other;
SQ
Query Match 47.6%; Score 1561.2; DB 7; Length 1707;
Best Local Similarity 95.0%; Pred. No. 0;
Matches 1624; Conservative 0; Mismatches 83; Indels 3; Gaps 1;
QY 66 ACGATGATGAACAAGCTTTAATGAGGAACTGAGCCCGCGCTACCGCGCAGACCT 125
Db 1 ACGATGATGAACAAGCTTTAATGAGGAACTGAGCCCGCGCTACCGCGCAGACCT 60
QY 126 CCGGAGCTCTTTGGGAGCAGAGAGTGCCTCGGCGGAGCAGAGTCTGTAAGTCCG 185
Db 61 CCGGAGCTCTTTGGGAGCAGAGAGTGCCTCGGCGGAGCAGAGTCTGTAAGTCCG 120
QY 186 CTAAGCCTCGTGAATAACCCGACCAAGAACTGGGCATCCGGCCATCGAGACCTCTC 245
Db 121 CTAAGCCTCGTGAATAACCCGACCAAGAACTGGGCATCCGGCCATCGAGACCTCTC 180
QY 246 GGGTAAAGTGAATTGCAATGGGAAATCATGGAATTGATTACTCACTCTTAAAGCT 305
Db 181 GGGTAAAGTGAATTGCAATGGGAAATCATGGAATTGATTACTCACTCTTAAAGCT 240
QY 306 AAGGACAGGAAATTCAGATTGAAACATCCCTCTCACTCACTGAGGAGGTGTTGA 365
Db 241 AAGGACAGGAAATTCAGATTGAAACATCCCTCTCACTCACTGAGGAGGTGTTGA 300
QY 366 TGACCTTTGGCTCAATATGAGCAGTGAAGATGTGAAACAATGCAACAACACAGA 425
Db 301 TGACCTTTGGCTCAATATGAGCAGTGAAGATGTGAAACAATGCAACAACACAGA 360
QY 426 AACCGCGTTGTCAGCTCAATATGCAACAAGAGAAAGCAAAAATAGCCATGAGAA 485
Db 361 GACCGCTGTTGTCAGCTCAATATGCAACAAGAGAAAGCAAAAATAGCCATGAGAA 420
QY 486 GCTAAGGGGCAATCAGTTGAGAACTACTCTTCAAGATTCTTCACTCCGAGTGAAG 545
Db 421 GCTAAGGGGCAATCAGTTGAGAACTACTCTTCAAGATTCTTCACTCCGAGTGAAG 480
QY 546 GGTAGCTCCCTTGGCCCTCGCCCTGACGAGCCAGCGTGGGAGCACTCTTCCGAGAGCA 605
Db 481 GGTAGCTCCCTTGGCCCTCGCCCTGACGAGCCAGCGTGGGAGCACTCTTCCGAGAGCA 540

```


QY AGGCCACGGCCCTGGGGGCACTTCTGAGCCGACAGATTGATTTTCCGCTCGGATCT 665
Db |||||
541 AGGCCAAGCCCTGGGGGCTCTTCTCAGGCCACAGAGATTGATTTTCCGCTCGGATCT 600
QY GGTCCCAACCCAGTTTGTGTGGCCATCTGCGAAAGAGGGCTTGACATTAAGACAT 725
Db |||||
601 GTTCCCAACCCAGTTTGTGTGGCCATCTGCGAAAGAGGGCTTGACATTAAGACAT 660
QY CACTAAGCAGACCAGTCCCGGGTAGATTCATGAAAGAGAACTCTGGAGCTGCAGCA 785
Db |||||
661 CACTAAGCAGACCAGTCCCGGGTAGATTCATGAAAGAGAACTCTGGAGCTGCAGCA 720
QY GAAAGCTGTCAACATCCATGCCACCCGAGAGGGACTTCTGAGCATGCCGATGATTTCT 845
Db |||||
721 GAAAGCTGTCAACATCCATGCCACCCGAGAGGGACTTCTGAGCATGCCGATGATTTCT 780
QY TGAATCATGCAAGAAAGAGGAGATGACCAAAATTACCCGAGAGATTTCTTGAAAT 905
Db |||||
781 TGAATCATGCAAGAAAGAGGAGATGACCAAAATTACCCGAGAGATTTCTTGAAAT 840
QY CTGGCACAATGGCTTGTGTGAAGACTGATTGAAAGAGGCAAAATTGAGAA 965
Db |||||
841 CTGGCACAATGGCTTGTGTGAAGACTGATTGAAAGAGGCAAAATTGAGAA 900
QY AATTGAATGAAAACAGGACCAAGATTAACATCTCATTTTGGAGATTGACATATA 1025
Db |||||
901 AATGAACATGAAAACAGGACCAAGATTAACATCTCATTTTGGAGATTGACATATA 960
QY CAACCCGAGAAAGAACCTCACTGTGAAGGGCAACGTTGAGGCTGTGCGACGTGAGAT 1085
Db |||||
1026 CAACCCGAGAAAGAACCTCACTGTGAAGGGCAACGTTGAGGCTGTGCGACGTGAGAT 1085
QY CAACCCGAGAAAGAACCTCACTGTGAAGGGCAACGTTGAGGCTGTGCGACGTGAGAT 1020
Db |||||
961 CAACCCGAGAAAGAACCTCACTGTGAAGGGCAACGTTGAGGCTGTGCGACGTGAGAT 1020
QY AGAGATTATGAAAGAGCTGCGTAGGCTTGAATAATGATGCTGGCTGTAAACACCA 1145
Db |||||
1086 AGAGATTATGAAAGAGCTGCGTAGGCTTGAATAATGATGCTGGCTGTAAACACCA 1145
QY AGAGATTATGAAAGAGCTGCGTAGGCTTGAATAATGATGCTGGCTGTAAACACCA 1080
Db |||||
1021 AGAGATTATGAAAGAGCTGCGTAGGCTTGAATAATGATGCTGGCTGTAAACACCA 1080
QY CTCGGATACTTCTCCAGCCTGTACCCCATCAACAGTTTGGCCGTTCCCGCATCATCA 1205
Db |||||
1146 CTCGGATACTTCTCCAGCCTGTACCCCATCAACAGTTTGGCCGTTCCCGCATCATCA 1205
QY CTTGCGATCTTCTCCAGCCTGTACCCCATCAACAGTTTGGCCGTTCCCGCATCATCA 1140
Db |||||
1081 CTTGCGATCTTCTCCAGCCTGTACCCCATCAACAGTTTGGCCGTTCCCGCATCATCA 1140
QY CTGTTATCCAGAGAGAGATTGTGATCTTCTTATCCCAACCGAGGCTGTGGCGGCAT 1265
Db |||||
1206 CTGTTATCCAGAGAGAGATTGTGATCTTCTTATCCCAACCGAGGCTGTGGCGGCAT 1265
QY CTGTTATCCAGAGAGAGATTGTGATCTTCTTATCCCAACCGAGGCTGTGGCGGCAT 1200
Db |||||
1141 CTGTTATCCAGAGAGAGATTGTGATCTTCTTATCCCAACCGAGGCTGTGGCGGCAT 1200
QY CATCGGAGAAAGGGGCAACATCAACAGCTGGCGAGATTGGCCGAGACCTTCATCAA 1325
Db |||||
1266 CATCGGAGAAAGGGGCAACATCAACAGCTGGCGAGATTGGCCGAGACCTTCATCAA 1325
QY CATCGGAGAAAGGGGCAACATCAACAGCTGGCGAGATTGGCCGAGACCTTCATCAA 1260
Db |||||
1201 CATCGGAGAAAGGGGCAACATCAACAGCTGGCGAGATTGGCCGAGACCTTCATCAA 1260
QY GATTGCCCCCTGGGAAAGGCCGAGAGCTCAGCGAAAGATGTCATCAACCGGCCACC 1385
Db |||||
1326 GATTGCCCCCTGGGAAAGGCCGAGAGCTCAGCGAAAGATGTCATCAACCGGCCACC 1385
QY GATCGCCCCCTGGGAAAGGCCGAGAGCTCAGCGAAAGATGTCATCAACCGGCCACC 1317
Db |||||
1261 GATCGCCCCCTGGGAAAGGCCGAGAGCTCAGCGAAAGATGTCATCAACCGGCCACC 1317
QY GGAAGCCCAAGTTAAAGCCCAAGGAGCGGATCTTTGGGAAACGTAAAGAGAACTTTCT 1445
Db |||||
1386 GGAAGCCCAAGTTAAAGCCCAAGGAGCGGATCTTTGGGAAACGTAAAGAGAACTTTCT 1445
QY GGAATCCCAAGTTAAAGCCCAAGGAGCGGATCTTTGGGAAACGTAAAGAGAACTTTCT 1377
Db |||||
1318 GGAATCCCAAGTTAAAGCCCAAGGAGCGGATCTTTGGGAAACGTAAAGAGAACTTTCT 1377
QY TAAACCCCAAGAGAGCTGAGAGCTGGAACCAATACAGAGTCCCTCTTCCACACTGG 1505
Db |||||
1446 TAAACCCCAAGAGAGCTGAGAGCTGGAACCAATACAGAGTCCCTCTTCCACACTGG 1505
QY TAAACCCCAAGAGAGCTGAGAGCTGGAACCAATACAGAGTCCCTCTTCCACACTGG 1437
Db |||||
1378 TAAACCCCAAGAGAGCTGAGAGCTGGAACCAATACAGAGTCCCTCTTCCACACTGG 1437
QY CCGGAGTATGGCAAAAGTGGCAAGACCGTGAACGAATTTAACTGATGTCAGATTAAT 1565
Db |||||
1506 CCGGAGTATGGCAAAAGTGGCAAGACCGTGAACGAATTTAACTGATGTCAGATTAAT 1565
QY CCGGAGTATGGCAAAAGTGGCAAGACCGTGAACGAATTTAACTGATGTCAGATTAAT 1497
Db |||||
1438 CCGGAGTATGGCAAAAGTGGCAAGACCGTGAACGAATTTAACTGATGTCAGATTAAT 1497
QY AGTCATCGGCTCCGTCGACCAACCGCATGAAATGAGAAATGATGTCAGATTAAT 1625
Db |||||
1566 AGTCATCGGCTCCGTCGACCAACCGCATGAAATGAGAAATGATGTCAGATTAAT 1625
QY AGTCATCGGCTCCGTCGACCAACCGCATGAAATGAGAAATGATGTCAGATTAAT 1557
Db |||||
1498 AGTCATCGGCTCCGTCGACCAACCGCATGAAATGAGAAATGATGTCAGATTAAT 1557
QY CCGGCACTTCTTGTGACGACGCTCAAGCGCAAGATCAGGGAATGTGACAAAGAT 1685
Db |||||
1626 CCGGCACTTCTTGTGACGACGCTCAAGCGCAAGATCAGGGAATGTGACAAAGAT 1685
QY CCGGCACTTCTTGTGACGACGCTCAAGCGCAAGATCAGGGAATGTGACAAAGAT 1617
Db |||||
1558 CCGGCACTTCTTGTGACGACGCTCAAGCGCAAGATCAGGGAATGTGACAAAGAT 1617
QY GAAAGCAGCAGAGCAGAAATACCTTCAGGGAAGTGGCTTCAGCGCAGCAAGTGAAGCTC 1745
Db |||||
1686 GAAAGCAGCAGAGCAGAAATACCTTCAGGGAAGTGGCTTCAGCGCAGCAAGTGAAGCTC 1745

Db |||||
1618 GAAAGCAGCAGAGCAGAAATACCTTCAGGGAAGTGGCTTCAGCGCAGCAAGTGAAGATC 1677
QY |||||
1746 CCACAGGCAACAGCAAAACAAACGATGAAT 1775
Db |||||
1678 CCACAGGCAACAGCAAAACAAACGATGAAT 1707
RESULT 11
AAS26570
ID AAS26570 standard, cDNA, 1186 BP.
XX
AC AAS26570;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human cDNA encoding a novel secreted protein, Seq ID 749.
KW
KW Human; immunosuppressive; antiarthritic; ss; antineumatic; cyostatic;
cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
antibacterial; virucide; fungicide; ophthalmological; vulnery;
secreted protein; rheumatoid arthritis; hyperproliferative disorder;
cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
cerebral ischaemia; angiogenesis; nervous system disorder;
Alzheimer's disease; infection; ocular disorder; corneal infection;
wound healing; epithelial cell proliferation; skin ageing; food additive;
preservative; antiproliferative.
KW
XX Homo sapiens.
OS
FN W0200155322-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001341.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 13-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0218390P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226682P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232089P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 25-SEP-2000; 2000US-0234999P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 08-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.

PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249219P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0251989P.
PR 06-DEC-2000; 2000US-0251479P.
PR 06-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI, 2001-488783/53.

P-PSDB; AAU16583.

PT New nucleic acid molecules encoding 461 human secreted proteins for
diagnosing, preventing, treating or ameliorating medical conditions and
used as food additives or preservatives.

Claim 1; SEQ ID NO 749; 980pp; English.

XX The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Antibodies to the proteins can also be used in
CC alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. cardiac
CC neoplasms of the breast or liver, cardiovascular disorders e.g. angioneu-
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angioneu-
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC and many other disorders listed in the specification. The polypeptides
CC can also be used to aid wound healing and epithelial cell proliferation,
CC to prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capability, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence encodes a novel secreted protein of the invention. Note: The
CC sequence data for this patent did not form part of the printed

Query Match 33.4%; Score 1096.8; DB 4; Length 1186;

Best Local Similarity 98.4%; Pred. No. 6.8e-281;

Matches 1145; Conservative 3; Mismatches 12; Indels 4; Gaps 4;

QY 1247 CCACGCTGTGAGCGCCATCATCGGGAAGAGGGGCAACATCAAAACAGCTGGCGAGAT 1306
DB 25 CCACGCGTCCGGGCATCATCGGGAAGAGGGGCAACATCAAAACAGCTGGCGAGAT 84

PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 17-NOV-2000; 2000US-0249239P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 XX (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX
 PI Rosen CA, Ruben SM, Barash SC;
 XX
 DR WPI; 2003-147444/14.
 DR P-PSDB; ABUS5651.
 XX
 PT New polypeptides and nucleic acids, useful in gene therapy for treating,
 PT inhibiting or preventing e.g. neural, immune system, muscular,
 PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
 PT renal disorders.
 XX
 PS Claim 1; SEQ ID NO 749; 402BP; English.
 XX
 CC The invention relates to human novel polypeptides and their associated
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene
 CC therapy for treating, inhibiting or preventing neural disorders, immune
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
 CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
 CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
 CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
 CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
 CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
 CC leukemia), inflammatory diseases (e.g. septic shock, bursitis and
 CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
 CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
 CC infarction) and cancerous diseases. Sequences ABX7173-ABX74167 represent
 CC human novel polynucleotides of the invention
 CC
 SQ Sequence 1186 BP; 338 A; 310 C; 302 G; 230 T; 0 U; 6 Other;
 Query Match 33.4%; Score 1096.8; DB 7; Length 1186;
 Best Local Similarity 98.4%; Pred. No. 6.8e-281;
 Matches 1145; Conservative 3; Mismatches 12; Indels 4; Gaps 4;
 QY 1247 CCCAGGCTGTGGCGCCCATCATCGGAGAGAGGGGCGACACATCAACAGCTGGCGAGAT 1306
 DB 25 CCCAGCGTTCGGCGCCCATCATCGGAGAGAGGGGCGACACATCAACAGCTGGCGAGAT 84
 QY 1307 TGGCGGAGCTTATCAAGATTGCCCTTGGGAGAGGCCCAAGCGTCAAGAGAGATGG 1366
 DB 85 TGGCGGAGCTTATCAAGATTGCCCTTGGGAGAGGCCCAAGCGTCAAGAGAGATGG 144
 QY 1367 TCATCAATCAACCGGGGCAACCGGAGCCCAAGTCAAGGCCGATCTTTGGAAAC 1426
 DB 145 TCATCAATCAACCGGGGCAACCGGAGCCCAAGTCAAGGCCGATCTTTGGAAAC 203
 QY 1427 TGAAGAAGGAAACTCTTTACCCCAAGAGAGAGTGAAGCTGGAAGCCATATCAAG 1486
 DB 204 TGAAGAAGGAAACTCTTTACCCCAAGAGAGAGTGAAGCTGGAAGCCATATCAAG 263
 QY 1487 TGGCCCTTTTCAAGCTGGCGGGGTGATTGGCAAAAGGTGGCAAGCCGTGAAGAACTGC 1546
 DB 264 TGGCCCTTTTCAAGCTGGCGGGGTGATTGGCAAAAGGTGGCAAGCCGTGAAGAACTGC 323
 QY 1547 AGAAGTTAACGAGTGAAGATCATGCTGCTCGTGAACCAAGCGCAGATGAATATGAG 1606
 DB 324 AGAAGTTAACGAGTGAAGATCATGCTGCTCGTGAACCAAGCGCAGATGAATATGAG 383
 QY 1607 AAGTGATCGTCAGAAATTAATCGGGCATCTTTGTGTAGCAGACTGACAGCGCAAGATCA 1666
 DB 384 AAGTGATCGTCAGAAATTAATCGGGCATCTTTGTGTAGCAGACTGACAGCGCAAGATCA 443

QY 1667 GGGAAATTG-TACAAACAGTGAAGCAGACAGAGAGAAATACCTTCAGGGAGTGGCTCA 1725
 DB 444 GGGAAATTGTTACAAACAGTGAAGCAGACAGAGAGAGAAATACCTTCAGGGAGTGGCTCA 503
 QY 1726 CAGCGCAGAAATGAGGCTCCCAACAGCACCAGCAAAACAGAGATGAATGAGCCCTTC 1785
 DB 504 CAGCGCA-NAAGTGAAGCTCCCAACAGCACCAGCAAAACAGAGATGAATGAGCCCTTC 562
 QY 1786 CAACACCTGACAGAAATGAGACCAACAGCAGCAGCAGATCGGAGAGCAACCAAGACCA 1845
 DB 563 CAACACCTGACAGAAATGAGACCAACAGCAGCAGCAGATCGGAGAGCAACCAAGACCA 622
 QY 1846 TCTGAGAAATGAGAAATGTGCGAGAGGCGGCGAGGACTCTGCGGAGGCTTGAAGACCC 1905
 DB 623 TCTGAGAAATGAGAAATGTGCGAGAGGCGGCGAGGACTCTGCGGAGGCTTGAAGACCC 682
 QY 1906 AGGGGCGGAGAGGGGCGGGGAGAGTCAAGGCTTGGCAAGCCAGAGCCCGGCT 1965
 DB 683 AGGGGCGGAGAGGGGCGGGGAGAGTCAAGGCTTGGCAAGCCAGAGCCCGGCT 742
 QY 1966 CCGGCCCCCGAGGGCTTCTGAGAGGCTTCAGGCAT-CCACTTCACCATCACTCGGATCTC 2024
 DB 743 CCGGCCCCCGAGGGCTTCTGAGAGGCTTCAGGCATTCACCATCACTCGGATCTC 802
 QY 2025 TCTGAACTCCCAAGAGCTATCCCTTTAGTTGAATPACTAGTGAAGCTTTCAA 2084
 DB 803 TCTGAACTCCCAAGAGCTATCCCTTTAGTTGAATPACTAGTGAAGCTTTCAA 862
 QY 2085 GCCAAGCAAAATGCAACCCCTTTCTGCGCAAAATGCTCTGTACATGTGTACATA 2144
 DB 863 GCCAAGCAAAATGCAACCCCTTTCTGCGCAAAATGCTCTGTACATGTGTACATA 922
 QY 2145 TTAGAAAGGAAAGATTAAGATATGAGGCTGTGGTTTACACAGGTCCTGACGGGT 2204
 DB 923 TTAGAAAGGAAAGATTAAGATATGAGGCTGTGGTTTACACAGGTCCTGACGGGT 982
 QY 2205 AATATATTTTAAAGAAATATATATCAATTAATCAATTAATCAATTAAT 2264
 DB 983 AATATATTTTAAAGAAATATATATCAATTAATCAATTAATCAATTAAT 1042
 QY 2265 TAATTTTTTTCTTTTAAAGAAAGAGGCTTTTCAAGCTTAAAGATTAAGTCT 2324
 DB 1043 TAATTTTTTTCTTTTAAAGAAAGAGGCTTTTCAAGCTTAAAGATTAAGTCT 1102
 QY 2325 TTGGAGAGCTCTACAGGTGTAGAGAGAGGCTTTGAGGCCACCGCAAAATTCACCA 2384
 DB 1103 TTGGAGAGCTCTACAGGTGTAGAGAGAGGCTTTGAGGCCACCGCAAAATTCACCA 1162
 QY 2385 GGGAAATCTGTCGGAAGGACT 2408
 DB 1163 GGGAAATCTGTCGGAAGGACT 1186
 RESULT 13
 AAC66035
 ID AAC66035 standard; cDNA; 1740 BP.
 XX
 AC AAC66035;
 XX
 DT 21-FEB-2001 (first entry)
 DE Human lung cancer-associated cDNA antigen L523S.
 XX Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
 KM vaccine; detection; ss.
 OS Homo sapiens.
 XX
 PN W0200061612-A2.
 XX
 PD 19-OCT-2000.
 XX

PF 03-APR-2000; 2000MO-US008896.
XX
PR 02-APR-1999; 99US-00285479.
PR 17-DEC-1999; 99US-00466396.
PR 30-DEC-1999; 99US-00476496.
PR 10-JAN-2000; 2000US-00480884.
PR 22-FEB-2000; 2000US-00510376.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Fan L,
XX
DR WPI, 2000-628399/60.
DR P-PSDB; AAB11365.
XX
PT Isolated polypeptide comprising an immunogenic portion of a lung tumor
PT protein is used for detecting and monitoring progression of lung cancer
PT in a patient.
XX
PS Claim 1a; Page 258-259; 261p; English.
XX
CC This invention describes a novel isolated polypeptide (I) which
CC comprising an immunogenic portion of a lung tumor protein or variant (P2)
CC which have cytostatic activity. The polypeptides and polynucleotides are
CC used in compositions and vaccines to inhibit the development of cancer,
CC especially lung cancer, in a patient. Methods described in the invention
CC can be used to monitor the progression of a cancer by carrying out the
CC detection at subsequent time points and comparing the results from the
CC different time points. CD4+ and/or CD8+ T-cells isolated from a patient
CC are treated with P2, polynucleotides encoding P2 or antigen presenting
CC cells expressing P2 and then administered to the patient to inhibit
CC development of cancer
XX
SQ Sequence 1740 BP; 526 A; 406 C; 417 G; 391 T; 0 U; 0 Other;

Query Match 21.2%; Score 697.2; DB 3; Length 1740;
Best Local Similarity 64.9%; Pred. No. 2.1e-174;
Matches 1116; Conservative 0; Mismatches 538; Indels 66; Gaps 3;

QY 73 ATGAAAGAGCTTTCATCGGGAAGCTGAGCCCGCCGTCACCGCCGACGACTTCGCGCAG 132
DB 1 ATGAAAGAGCTTTCATCGGGAAGCTGAGCCCGCCGTCACCGCCGACGACTTCGCGCAG 60

QY 133 CTCTTTGGGGACAGAGAGCTGCCCCCTGGCGGGACAGGTCCTGTGAAGTCCGGGTAAAGCC 192
DB 61 ATCTTCAAGACGCGCAAGATCCCGGTGTGCGGACCCCTTCGTGGAAGACTGGCTACGCG 120

QY 193 TTGCTGACCTACCCCGACAGAACTGCGCCATTCGCGCCATCGAAGCCCTCTCGGCTAAA 252
DB 121 TTCTGTGACCTGCCCGGACGAGAGCTGGGCCCTCAAGGCCATCGAGCGCTTTCAGGTAAA 180

QY 253 GTGGAATTGCATGGGAAATCATGGAAGTTGATTATCATCTCTTAAAGCTTAAGAGC 312
DB 181 ATGAAATCTGACGCGAAACCCATAGAAAGTTGAGCAGCTGCTCCAAAAGCAAGATT 240

QY 313 AGGAAATTCAGATTGGAACATCCCTCTCACTGAGTGGGAGGTGGTGAATGACT 372
DB 241 CGGAACTTCAGATTGGAATATCCCTCTCACTTAAAGTGGAGGTGCTGATAGTTTA 300

QY 373 TTGGCTCAATATGAGGACATGGAAGATGTGAAACAACTCAACAGACAGAAACCGCC 432
DB 301 CTAGTCCAGATAGAGAGTGTGAGAGCTGTGAGCACTGACTGCTGGAACCTGCA 360

QY 433 GTTGTCAAGCTCATATTTGCAACAAGAGAAAGCAAAATAGCCATGGAAGCTAAGC 492
DB 361 GTTGTAAATGTAATCTATTCAGATTAGAGCAAGCTAAGACAACTGAAT 420

QY 493 GGGCATCAGTTTGAAGACTCTCTCAAGATTCTCTACATCCCGATGAAGAGGTGAGC 552
DB 421 GATTTCAAGTTAAGAAATTCACCTTGAAGAGTACCTATTCCTGTATGAACGCGCGCT 480

QY 553 TCCCTTCGCGCCCTCAGGAGCCCAAGCTGGGGAACAATCTTTCGCGGAGCAAGGC --- 609

DB 481 CAGCAAAACCCCTTGACAGAGCCCGAGAGTGCCTGGGGCTTGGGACAGAGGCTCTCA 540
QY 610 -----CAGCGCCCTGGGGGACCTTCTCAGGCGACAGATTTATTTCCCGTGGGATC 663
DB 541 AGGCAGGGGTCTCCAGGATCCGTAATCAAGGAGAAACATGATTTGCTCTTGCGCGCTG 600
QY 664 CTGGTCCACCCATTTGTTGTGTGTCATCATCGAAAGAGAGGCTTGACCAATTAAGAAC 723
DB 601 CTGTTCCACCCAAATTTGTGAGCCATCATAGGAAAGAAAGTGGCCACCATTTGGAAAC 660

QY 724 ATCACTAAGCAGACCCAGCTCCCGGTAGATATTCATAGAAAGAAAGAACTGTGAGGTGA 783
DB 661 ATCAACCAACAGACCCAGCTTAAATCATGTGTCACCGTAAAGAAATGCGGGGCTGT 720

QY 784 GAGAAAGCTGTACCATCATGCAACCCAGAGGGGACCTTGAAAGCATGCGCATATT 843
DB 721 GAGAAAGTGAATTAATCTCTCTCTCACTCTGAAAGCACTTCGCGGCTTGAAGTCTATT 780

QY 844 CTGGAATTCATGCGAAGAAAGAGGCAATGAGACAAACTAAGCCGAAGATTCCTGAAA 903
DB 781 CTGGAATTCATGCGAAGAAAGAGGCAATGAGATTAATTAATTCACAGAAAGATCCCTTGAAG 840

QY 904 ATCTTGGCACAAATGCTGTGTTGAAGATGGAAGTGAAGAAAGAGCAAAATTTGAAG 963
DB 841 ATTTTAGCTAATTAATCTTTGTTGAGCTCTTATTGTAAAGAAAGAAATCTTAAA 900

QY 964 AAAATGAAACATGAAACAGAGGACCAAGTAAACAATCTCATCTTTGACAGATTTGACATA 1023
DB 901 AAAATGAGCAAGACAGACAGACATTAATATCGAATATTCATTTGACAGAAATTTGACGCTG 960

QY 1024 TACAAACCGGAAAGAAACATCATCTGTAAGGACAGCTTGAAGGCTGTGCGACGTGAG 1083
DB 961 TATATCCGAACCCACTATTAATCAATTAAAGCAATGTTGAGACATGTCGCAAGAGCTGAG 1020

QY 1084 ATAGAGATTAATGAAGAGCTGCGTGAAGCCCTTGAAATATGATGCTGCTTAAACAC 1143
DB 1021 GAGAGATTCATGAAGAAATATGAGGAGTCTTATGAATGATTTGCTTATGATATCTT 1080

QY 1144 CACTCGGAT-----ACTTCCCA 1162
DB 1081 CAAGCATTTAATTCCTGATTAATCTGAAGCCCTTGGTCTGTTCCACCACTTCA 1140

QY 1163 GCGTGTACCCCATCACCAGTTTGGCCGTTCCCGCATCATCTC----- 1208
DB 1141 GGGATGCACTCCCACTCAGGCGCCCTTCAGCCATGACTCCTCCCTAACCCGAGTTT 1200

QY 1209 --TTATCAGAGCAGAGATTTGTAATCTCTCAATCCCAACCCAGGCTGTGGGGCATC 1266
DB 1201 GAGGAATCAGAAAGAGAGCTGTTCACTGTGTTATCCAGCTCTATACAGTGGGCAATC 1260

QY 1267 ATGGGAAAGAGGGGCAACATCAACAGCTGGCGAATTCGCGGAGCTCTATCAAG 1326
DB 1261 ATCGGCAAGCAGGGCCAGCATCAAGCACTTCTCGCTTGTGAGAGCTTCAATTAAG 1320

QY 1327 ATTGCCCTGCGGAAGGCCCCAGAGCTCAGGAAAGATGTGTCATCATCCGCGCACCG 1386
DB 1321 ATTGCTCAAGCGAAGACCAAGTGTCTAAAGTGAAGTGTGATTAATCACTGACACACA 1380

QY 1387 GAAGCCAGTTCAAGGCCGAGAGCGGATTTTGGGAAATGAAAGAGAAATCTTT 1446
DB 1381 GAGGCTCAGTTCAAGGCTCAGGGAAGAAATTTATGAAAGAAATTAAGAAAGAAATCTTTGTT 1440

QY 1447 AACCCAAAGAAAGATGAAAGCTGAAGGCGCATATGAGATGCGCTCTTCAAGAGTGC 1506
DB 1441 AGCTCTAAAGAAAGAGGTGAAGCTTGAAGCTCATATGAGATGCTCATCTTTGTGCTGCGC 1500

QY 1507 CGGGTATTTGGCAAAAGTGGCAAGACCGTGAAGAACTGAGAACTTAACAGTGCAGAA 1566
DB 1501 AGAGTTATTGGAAGAAAGAGGCAAAAGGTGAAGAACTTCAAGATTTGTCAAGTGCAGAA 1560

QY 1567 GTTCATGCTGCTGTGACCAACAGCGCAGATGAAATGAAGAGATGATCGTCAATTAATC 1626
DB 1561 GTTGTGTCTCTGTGACAGACACTGATGAGATGACCAAGTGGTTGTCAAAATTAATCT 1620

[illegible]

RESULT 14

ID	ABL49254 standard; cDNA; 1740 BP.
XX	
AC	ABL49254;
XX	
DT	01-MAY-2002 (first entry)
XX	
DE	Human lung tumour L523S cDNA sequence SEQ ID NO:347.
XX	
KW	Human; lung tumour; lung cancer; cytostatic; immunostimulant; vaccine; immune response; ss.
KW	

PA (CORI-) CORIXA CORP.

PI Wang T, Wang A, Sheiky YAM, Li SX, Kalos MD, Henderson RA,
PI McNeill PD, Fanger N, Retter MW, Marnerakis M, Fanger GR,
PI Vedvick TS, Carter D, Watanabe Y, Peckham DW;
XX
DR MPI: 2002-090513/12.
P-PSDB; ABB74997.
DR

PT polynucleotides encoding lung tumor polypeptides, useful for treating lung cancer or stimulating an immune response.

PS Example 2; Page 330; 374pp; English.

The present invention describes human lung tumour proteins. Human lung tumour proteins and polynucleotides have cytostatic and immunostimulant activities, and can be used in vaccine production. Compositions comprising the lung tumour proteins, polynucleotides, antibodies, fusion proteins, T cell populations, or antigen presenting cells that express the lung tumour proteins are useful for treating lung cancer or stimulating an immune response. ABL4895 to ABL4300 and ABB74946 to ABB75070 represent sequences used in the exemplification of the present invention.

SQ Sequence 1740 BP; 526 A; 406 C; 417 G; 391 T; 0 U; 0 Other;

Query Match	21.2%	Score 697.2	DB 6	Length 1740
Best Local Similarity	64.9%	Pred. No. 2.1e-174		
Matches 116; Conservative	0	Mismatches 538	Indels 66	Gaps 3

Qy 73 ATGAAACAGCCTTACATCGGGAACCTGAGCCCGCGCTACCGCGAGCAG 13
Db 1 ATGAAACAACTGTATATCGAAACCTCAGCAGAACCGCCGCCCTCGAAGCTGAAAGT 60

QY	133	CTCTTTGGGACAGAGAACTGCCCTCGGGGGGACAGGCTCTGTAACTCCGACTACGCC	192
Db	61	ATCTCTTAAGAGACGCCAAGATCCCGGGTGTGGGACCTCTCTGGTAAAGACTGGCTACCGG	120
QY	193	TTTGCTGACATACCCCGACAGAACTGGGCGCATCCGGCCCATCGAAGCCCTCTCGGGTAAA	252
Db	121	TTTGCTGACATCGCCCGGACGAGAGCTGGGCGCTTAAGGCGCATCGAGCGCTTTACAGTAAA	180
QY	253	GTCGAATTGCATGGGAAAAATCATGGAGTTGATTACTAGTCTCTAAAAAGTTAAGAGGC	312
Db	181	ATAGAACTGCACGGGAAACCCATAGAAAGTTGAGCACTCGCTCCAAAAAGCAAGGATT	240
QY	313	AGGAAATTTACAGTTTGAAACATCCCTCTCACTCGCAGTGGGAGGTGTGGATGACTT	372
Db	241	CGGAAACTTACAGTATCGAAATATCCCGCTCATTTTACAGTGGGAGGTGTGGATGATT	300
QY	373	TTTGCTCAATATATGGGACAGTGGAGAAATGTGGAAACAATGACACAGACACAGAAACCGCC	432
Db	301	CTAGTCCAGATGAGAGTGGTGGAGACCTGTGAGCAAGTACACTGACCTCGGAAACTGCA	360
QY	433	GTTGTCAACGTCACATATGCAACAAGAGAAAGAAAGCAAAATATGCCATGGAGAAAGTTAAGC	492
Db	361	GTTGTAAATATGAACTATTCAGATAGAGACCAAGCTAGCAAGCACTAACAACACTGAT	420
QY	493	GGGATCATGTTTGAAGACTACTCTCTCAAGATTTCTCTCATCCCGGATTAAGAGGTAGC	552
Db	421	GGATTTAGTTAGTAAGAAATTTCACTTGBAAGTBACTTATCTCTGATTAAGACGGCCGC	480
QY	553	TCCCTCTCGCCCCCTCAGCAGAGCCACAGCTGGGGACCACTCTTCCGGAGCAAGGC---	609
Db	481	CAGCAAAACCCCTTGCAGAGAGCCCGGAGGTGGCGGGGCTTGGCAGAGGGGCTCTCTCA	540
QY	610	-----CAGCGCCCTG9GGG9CACTTCTCAGGCGCAGACAGATGATTTTCCCGCTGCGGATC	663
Db	541	AGCAGAGGGGTCTCAGAGATCCGATCCAGACAGAAACCATGTGATTTGCTCTGCGCCTG	600
QY	664	CTGCTCCCAACCCAGTGTGTTGGGCATATGCGAAGAGAGGGGCTTGACCATATAAGAAC	723
Db	601	CTGGTCTCCACCAATTTGTTGGAGCCATATAGAAAGAAAGAGTGGCACACTTCGGAAAC	660
QY	724	ATCATTAAGCAGACCCAGTCCCGGGTATGATATCATAGAAAGAGAACTCTGGAGCTGCA	783
Db	661	ATCACAACACAGACCCAGTCTTAAATATGATGTCAACCGTATAAGAAATGCGGGGCTGCT	720
QY	784	GAGAAAGCTGTACCATTCATGCCACCCGAGGGGACTTCTGAACATAGCCGCGCATGATT	843
Db	721	GAGAAAGTGAATTAATCTATCTCTACTCTCGAAGGCACTCTGCGGTGTAAAGTCTATT	780
QY	844	CTTGAATATCAGCAGAAAGAGCAGATGAGACCAACCTAGCCGAAAGATTTCTCTGAAA	903
Db	781	CTGGAGATTAATGCACTAAGAGGCTCAAGATATTAATTCACAGAGAGATCCCTTGAAAG	840
QY	904	ATCTTGGACACAATGCGCTGTGTGGAAAGACTGATTTGAAAAGAAAGGACAGAAATTTGAAG	963
Db	841	ATTTTAGCTCATATTAATCTTGTGGAGCGTCTTATTTGTAAAGAAAGAAATCTTAAA	900
QY	964	AAAAATTGAACATGAACAGGACCAAGATTAACAATCTCAATCTTTGCAAGATTTGAGATA	1023
Db	901	AAAAATTGAGCAAGACACAGACACTAAATACGATATCTCATTTGACAGGATTTGACGCTG	960
QY	1024	TACAAACCCGGAAGAACATCACTGTGAAGGGGACAGTTGAGGCTGTGCACAGTCTGAG	1083
Db	961	TATTAATCCAGAAACGCAATATTAACGATTAAAGGCAATGTTGAGACATGTGCCAAAGCTGAG	1020
QY	1084	ATAGAGATTAATGAAGAGCTGCGTAGGCGCTTTGAAAAATGATATGCTGGCTGTTAACCC	1143
Db	1021	GAGGAGATATATGAAGAAATATCAGGAGCTTATGAATATGATTTGCTTATGAATCTT	1080
QY	1144	CACTCGGAT-----ACTTCTCCA	1162
Db	1081	CAAGCAATTAATTCCTGATTTAATCTGAACGCTTGGGTCTGTTCCACCACTTCA	1140
QY	1163	GCGCTGATACCCCATACAGATTTGGCGCGGTTCCGCGCATATCACTC-----	1208

Db 1141 GGGATGCCACCTCCACCTCAGGCGCCCTTCAGCCATGACTCTCCCTACCCGAGTTT 1200
 Qy 1209 --TTATCCAGAGAGAGATGATGATCTTCATCCCAACCAGGCTGTGGCCGATC 1266
 Db 1201 GAGCATCAAGAACGAGACTGTCTATCTGTTATCCCAAGCTCTATCACTGATGATC 1260
 Qy 1267 ATCGGGAAGAAAGGAGGAGCAGATCAAAACAGCTGGGAGATTGGCCGAGCCCTTATCAAG 1326
 Db 1261 ATCGGCAAGAGGCGCAGACATCAAGAGAGCTTTCTGCTTGTGGAGCTTCAATTAG 1320
 Qy 1327 ATTCGCTCGGAGAGGCGCAGAGCTGACGGAAGATGTCATCATCAACGCGGACACG 1386
 Db 1321 ATTCGCTCGGAGAGCAGAGCTGATTAAGTGAAGATGATGATATCATCTGAGACCA 1380
 Qy 1387 GAAGCCAGTTCAAGGCGCAGGAGAGATCTTTGGGAAATGAAAGAGAAATCTTTT 1446
 Db 1381 GAGGCTCAGTCAAGGCTCAGGAGAAATTTATGAAAAATTTAAAGAAAACTTTGTT 1440
 Qy 1447 AACCCAAAGAAAGATGAGCTGAGAGCGCATATCAGAGTGCCTTCCACAGCTGAC 1506
 Db 1441 AGTCTTAAGAAAGAGTGAATTAATTAAGTGAATGATGATGATGATGATGATGATG 1500
 Qy 1507 CGGCTGATGGGAAAGTGTGCAAGACGCTGAACGAACTGAGAACTTAACGATGACAG 1566
 Db 1501 AGAGTTATTTGAAAAAGAGAGCAAAACGCTGAATGAATCTTCAAGATTGTCAAGTG 1560
 Qy 1567 GTCAATCGCTCGGAGCAACGCGCATGATGAATGAGAAAGTATGATGATGATGATGATG 1626
 Db 1561 GTTGTGTCTCTGCTGAGCAGACCTGATGAGATGACCAAGTGTGTGCAAAATTAAT 1620
 Qy 1627 GGGCACTTCTTTGCTAGGAGAGCTGAGAGCGCAAGATCAGGAAATTTGTACAAAGG 1686
 Db 1621 GGTACTTCTATGATGCTGAGAGGTTGCCAGAGAAATTTCTGAGTCAAGGTA 1680
 Qy 1687 AAGCAGAGAGAGCAAAATACCTCAGGAGTGCCTCAG 1726
 Db 1681 AAGCAGAGAGCAAAAGAGCTCTGCAAAAGTGCAGCAC 1720
 RESULT 15
 ID AB092440 standard; cDNA; 1740 BP.
 AC AB092440;
 XX
 DT 07-OCT-2002 (first entry)
 DE Human lung cancer associated cDNA sequence SEQ ID NO:347.
 XX
 KW Human; lung cancer; lung tumor; cytostatic; gene therapy; vaccine; gene;
 KM ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200247534-A2.
 PD 20-JUN-2002.
 XX
 PF 30-NOV-2001; 2001MO-US047576.
 XX
 PR 12-DEC-2000; 2000US-00735705.
 PR 07-MAY-2001; 2001US-00850716.
 PR 28-JUN-2001; 2001US-00897778.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Wang T, Wang A, Skeiky YAM, Li SX, Kalos MD, Henderson RA;
 PI Monelli PD, Fanger N, Retter MW, Durham M, Fanger GR, Vedrick TS;
 PI Carter D, Watanabe Y, Peckham DW, Cai F, Foy TM;
 XX
 DR WPI; 2002-583465/62.
 DR P-FSDB; ABP61917.

XX Novel lung carcinoma polynucleotide sequences and polypeptides encoded by
 PT the polynucleotides, useful in pharmaceutical compositions such as
 PT vaccines and as markers to indicate the presence of lung cancer.
 PS Claim 1, Page 337; 381pp; English.
 XX
 CC The present invention describes isolated human lung carcinoma
 CC polynucleotides (I) and polypeptides (II). (I) and (II) have cytostatic
 CC activity, and can be used in gene therapy and in vaccines. Compositions
 CC comprising (I) or (II) can be used for stimulating an immune response in
 CC a patient and for treating lung cancer in a patient. Oligonucleotides of
 CC (I) can be used for detecting the presence of a cancer in a patient, by
 CC obtaining a biological sample from the patient, contacting the biological
 CC sample with the oligonucleotide, detecting in the sample, an amount of
 CC polynucleotide that hybridizes to the oligonucleotide and comparing the
 CC amount of polynucleotide that hybridizes to the oligonucleotide to a
 CC predetermined cut-off value, and determining the presence of a cancer in
 CC the patient. (I) and (II) are useful in pharmaceutical compositions, e.g.,
 CC vaccines. (I) is useful as a marker to indicate the presence or absence
 CC of a cancer such as lung cancer. AB092445 to AB092486 and ABP61866 to
 CC ABP61992 represent sequences used in the exemplification of the present
 CC invention
 CC
 SQ Sequence 1740 BP; 526 A; 406 C; 417 G; 391 T; 0 U; 0 Other;

Query Match 21.2%; Score 697.2; DB 6; Length 1740;

Best Local Similarity 64.9%; Pred. No. 2.1e-174;

Matches 116; Conservative 0; Mismatches 538; Indels 66; Gaps 3;

Qy 73 ATGAACAAGCTTATCATTCGGAACCTGAGCCCGCGTACCGCGACCTCCGAG 132
 Db 1 ATGAACAAGCTTATCATTCGGAACCTGAGAGAACCGCCCGCTGACCTTGAAGT 60
 Qy 133 CTCTTTGGGAGCAGAGAGCTGCCCCGCGGAGAGATCCCTGAGTCCGCGTACGCC 192
 Db 61 ATCTTCAAGAGAGCGCAAGATCCCGTGTGCGAGCCTTCTGTAAGAGCTGCTACCG 120
 Qy 193 TTGCTGAGCTACCCCGACAGAACTGGGCGCATCCGCGCATCGAACCTTCCGGTAA 252
 Db 121 TTGCTGAGCTACCCCGACAGAGAGTGGGCGCATCGAGCGGCTTTCAGGTA 180
 Qy 253 GTGAATTCATGAGAAATCATGGAAGTGAATTCATGATCTCTTAAAGCTAAGAGC 312
 Db 181 ATGAAGCTGACCGGAAACCCATAGAACTGAGACCTCGTCCCAAAAGCAAGATT 240
 Qy 313 AGAAAAATTCAGATTGGAATTCATCCCTCCTCAGCTGAGGAGGAGTGTGAGACTT 372
 Db 241 CGAAATTCATGATGAGAAATATCCGCTCTATTAAGTGGAGGAGTGTGATGTTTA 300
 Qy 373 TTGCTCATATATGAGAGCAGTGAAGATGTGGAACAAGTCAACAGACAGAAACCGCC 432
 Db 301 CTAGTCAGATATGAGAGTGTGAGAGAGCTGAGCAAGTGAACACTGACTGGAACTGCA 360
 Qy 433 GTTGTCAAGCTCATATGTCACAAGAGAAAGCAAAATATGACTGAGAGAACTAAGC 492
 Db 361 GTTGTAAATGTAACCTATTCAGTAAGACCAAGCTAGACAACTAGACAAACTGAAT 420
 Qy 493 GGGCATCAGTTTGAAGACTATCTCTCAAGATTCTCAGATCCCGGATGGAAGGTAGC 552
 Db 421 GGAATTCAGTTAAGAAATTTCACTTGAAGAGTATGCTTATGTAAGAGCGCGCC 480
 Qy 553 TCCCTTCGCCCCCTCAGAGAGCCAGCGTGGGAGCACTCTTCCGAGAGCAAGG--- 609
 Db 481 CAGCAAAACCCCTTTGAGAGAGCCCGAGGTCGCGGGGCTTGGGAGAGGAGGCTCTCA 540
 Qy 610 -----CAGGCCCTTGGGGGCACTTCTCAGGCGACAGATGATTTCCGCTGGAGTC 663
 Db 541 AGGAGAGGAGTCTTCAGAGATCCGATCAAGCAAGAACTATGATTTGCTTCGCGCTG 600
 Qy 664 CTGATCCCAAGCAGATTTGTGCTGATCATCGAAAGAGAGGCTTGAACATTAAGAGC 723
 Db 601 CTGATCCCAAGCAGATTTGTGAGGCACTATGAGAAAGAGGTGCCACATTCGGAAC 660

QY 724 ATCACTAAGCAGACCCAGTCCCGGTAGATATCCATAGAAAAGAACTCTGGAGCTGCA 783
Db 661 ATCAACCAAAACAGACCCAGTCTAAATCGATGCTCAACCTTAAAGAAAAGCGGGCGCTGCT 720
QY 784 GAGAAAGCTGTCACCATCATGCCCCAGAGGGGAGCTTCTGAAGCATGCCGATGAT 843
Db 721 GAGAAAGTGAATTAATCTCTCTTACTCTTGAAGGACCTCTGGCGCTTGTAAAGTAT 780
QY 844 CTGGAATCATGCAAGAAAAGGAGCAGATGAGACCAACTAGCCGAAGAGATTCTTGAAA 903
Db 781 CTGGAGATTAAGCAATAGAAAGCTCAAGATATAAATTCACAGAAAGATCCCCCTTGAAG 840
QY 904 ATCTTGGCAGCATGCTGCTGTTGGTTGGAAGCTGATTGGAAGAAAGAGCAAAATTGAA 963
Db 841 ATTTAGCTCAATAATTAATCTTTGAGCGTCTTATTTGTAAGAAAGAAATCTTAA 900
QY 964 AAAATTGAACATGAACAGAGGACCAAGATTAACAATCTCATCTTTGCAGAGATTGACATA 1023
Db 901 AAAATTGAGCAGACACAGACACTAAATCAAGATATCTCCATTCAGAGAAATTGACGCTG 960
QY 1024 TACAAACCCGAAAAGAACATCACTGTGAAGGGCAAGTTAGGGCTGTGCCAGTGTAG 1083
Db 961 TATAATCCAGAACGCACTATTACGTTAAAGCAATGTTGAGACATGTGCCAAAGCTGAG 1020
QY 1084 ATAGGATTAATGAAGAAGTGTGCTGAGGCTTTGAAAATGATATGCTGCTTAACACC 1143
Db 1021 GAGGAGATCATGAAGAAATCAGGAGTCTTATGAATAATGATTTGCTTTATGAACTTT 1080
QY 1144 CACTCCGGAT-----ACTTCTCCA 1162
Db 1081 CAAGACATTTAATCTCTGATTAATCTGAAGGCTTGGCTGTGTTCCCACTTCA 1140
QY 1163 GCCTGTACCCCATCAACAGTTTGGCCGCTTCCCGATCATCTC----- 1208
Db 1141 GGGATGCCACTCCCACTCAGGGCCCTTCAGCCATGACTCTCCCTACCCGCAATT 1200
QY 1209 --TTATCCAGACAGAGATTGTGAATCTTCATCCCAACCAAGCTGTGGGGCCCATC 1266
Db 1201 GAGCATCAAGAAAGAGACTGTCTATCTGTTATCCAGCTCTATCAGTGGTGCAATC 1260
QY 1267 ATCGGAGAAAGAGGGGACACATCAAAACAGCTGCGAGATTGCGCGAGCTCTATCAAG 1326
Db 1261 ATCGGAAAGAGGGCCAGACACATCAAGCACTTTCGCTTGTGAGCTTCAATTAAG 1320
QY 1327 ATTGCCCTGCGAAGGCCCAAGCTCAGCGAAAGAGTGTCTATCAACCGGCCACCG 1386
Db 1321 ATTGCTCCAGCGAAGCAACAGATGCTAAAGTGAGATGTGATATCACTGCAACCA 1380
QY 1387 GAAGCCCACTTCAAGGCCCAAGGACGATCTTTGGGAACTGGAAGAGAAAACCTCTTT 1446
Db 1381 GAGGCTCAGTTCAGGCTCAGGGAAGATTTATGAAAATTAAGAAAGAAACCTTTGTT 1440
QY 1447 AACCCCAAGAAAGATGAAGCTGAAGGCGCATATCAGAGTGCCTCTTCCACAGCTGGC 1506
Db 1441 AGTCTTAAAGAGAGGTGAAACTTGAAGCTCATATCAGAGTGCATCTTGTGCTGGC 1500
QY 1507 CCGGTATTTGGCAAGGTGGCAAGCCGTGAACGAACCTGCAAACTTAACCACTGACAGA 1566
Db 1501 AGAGTTATTGAAAAGAGGCAAAAGGTGAATGAACCTTCAGAAATTTGTCAAGTCAGAA 1560
QY 1567 GTATCTGTGCTGTGATCCAAACGCCAGATGAATAATGAGAAATGATCGTCAAGATTATC 1626
Db 1561 GTTGTGTCTCTGTGTGACAGACCTGATGAGATGACCAAGTGTGTCAAAATTAAT 1620
QY 1627 GGGCACTTTCTTGTAGCCAGACTGACAGCGCAAGATCAGGAAATTGTACAAACAGGTG 1686
Db 1621 GGTCACTTCTATGCTTGTCCAGTGTGCCAAGAAAATTCAGGAAATTCGACTCAGGTA 1680
QY 1687 AAGCAGAGAGAGAAATACCTTCAGGGAGTGCTCAC 1726
Db 1681 AAGCAGACCAACAGAGAGGCTCTGCAAGTGACCAAC 1720

Search completed: July 24, 2004, 07:28:53
Job time : 1231.37 secs

Db 301 AAGCTAAGAGAGGAAAAATTCAGATTGAAAACATCCCTCTCACTGTGAGTGGAGGTG 360
Qy 361 TTGATGGACTTTTGGCTCAATATATGAGACATGAGAAATGTGGAAACAATCAACAGAC 420
Db 361 TTGATGGACTTTTGGCTCAATATATGAGACATGAGAAATGTGGAAACAATCAACAGAC 420
Qy 421 ACGAAACCGCGGTGTCAAGTCACATATGCAACAGAAAGAAAGCAAAAATAGCCATG 480
Db 421 ACGAAACCGCGGTGTCAAGTCACATATGCAACAGAAAGAAAGCAAAAATAGCCATG 480
Qy 481 GAGAAGCTAAGCGGCGATGTTTGAGAACTACTCTCTTCAGATTTCCTACATCCCGAT 540
Db 481 GAGAAGCTAAGCGGCGATGTTTGAGAACTACTCTCTTCAGATTTCCTACATCCCGAT 540
Qy 541 GAAAGGTGAGCTCCCTTGCGCCCTCAAGCAGCCCAAGTGGGAGCACAATCTCCCG 600
Db 541 GAAAGGTGAGCTCCCTTGCGCCCTCAAGCAGCCCAAGTGGGAGCACAATCTCCCG 600
Qy 601 GAGCAAGGCGAGCGCCCTGGGGGCACTTCTCAGGCGCAGACAGATTATTTCCGCTCGG 660
Db 601 GAGCAAGGCGAGCGCCCTGGGGGCACTTCTCAGGCGCAGACAGATTATTTCCGCTCGG 660
Qy 661 ATCTGTGCTCCCAACCCAGTTTGTGTGTCATCATCGAAAGAGGCTTGAACATTAAG 720
Db 661 ATCTGTGCTCCCAACCCAGTTTGTGTGTCATCATCGAAAGAGGCTTGAACATTAAG 720
Qy 721 AACATCACTAAGACAGCCAGTCCCGGGTATGATATCATAGAAAAGAACTCTGAGCT 780
Db 721 AACATCACTAAGACAGCCAGTCCCGGGTATGATATCATAGAAAAGAACTCTGAGCT 780
Qy 781 GAGAGAAAGCTGTACCATTCATGCAACCCAGAGGGGCTTCTGAGCATGCGCATG 840
Db 781 GAGAGAAAGCTGTACCATTCATGCAACCCAGAGGGGCTTCTGAGCATGCGCATG 840
Qy 841 ATTCTTGAATCATGAGAAAGAGGAGATGAGACCAAACTAGCCGAGAGATTCTCTG 900
Db 841 ATTCTTGAATCATGAGAAAGAGGAGATGAGACCAAACTAGCCGAGAGATTCTCTG 900
Qy 901 AAAATCTTGGCACACATGCGCTTGTGGAAAGCTGATTTGAAAAGAAAGCAAAAATTTG 960
Db 901 AAAATCTTGGCACACATGCGCTTGTGGAAAGCTGATTTGAAAAGAAAGCAAAAATTTG 960
Qy 961 AAGAAATTAAGATGAACAGGAGCCAGATTAACATCTCATCTTTCAGATTTGAGC 1020
Db 961 AAGAAATTAAGATGAACAGGAGCCAGATTAACATCTCATCTTTCAGATTTGAGC 1020
Qy 1021 AATTAACAACCCGAAAGAACCATCACTGTGAAAGGCGACAGTTGAGGCTTGCAGTGT 1080
Db 1021 AATTAACAACCCGAAAGAACCATCACTGTGAAAGGCGACAGTTGAGGCTTGCAGTGT 1080
Qy 1081 GAGATAGAGATTATGAAAGAGCTGCGTGAAGCCCTTTGAAAATGATATGCTGTGTAAC 1140
Db 1081 GAGATAGAGATTATGAAAGAGCTGCGTGAAGCCCTTTGAAAATGATATGCTGTGTAAC 1140
Qy 1141 ACCCACTCGGGAATCTTCTCAGCTGTACCCCATACAGTTTGGCCCGCTTCCGCAT 1200
Db 1141 ACCCACTCGGGAATCTTCTCAGCTGTACCCCATACAGTTTGGCCCGCTTCCGCAT 1200
Qy 1201 CATCACTCTTATCCAGAGAGAGATTGTGAATCTTTCATCCCAACCCAGGCTGTGGC 1260
Db 1201 CATCACTCTTATCCAGAGAGAGATTGTGAATCTTTCATCCCAACCCAGGCTGTGGC 1260
Qy 1261 GCCATCATCGGGAAGAGGCGGCGACATCAAAAGCTGCGAGATTGCGCGAGCTCT 1320
Db 1261 GCCATCATCGGGAAGAGGCGGCGACATCAAAAGCTGCGAGATTGCGCGAGCTCT 1320
Qy 1321 ATCAAGATTGCGCTGCGGAAGGCGCAGAGTCAAGCAAAAGATGTCATCATACCGGG 1380
Db 1321 ATCAAGATTGCGCTGCGGAAGGCGCAGAGTCAAGCAAAAGATGTCATCATACCGGG 1380
Qy 1381 CCAACCGGAAGCCAGTTCAAGGCGCCAGGAGCGGATCTTTGGAAAAGTGAAGGAAAAC 1440
Db 1381 CCAACCGGAAGCCAGTTCAAGGCGCCAGGAGCGGATCTTTGGAAAAGTGAAGGAAAAC 1440

Db 1381 CCAACCGGAAGCCAGTTCAAGGCGCCAGGAGCGGATCTTTGGAAAAGTGAAGGAAAAC 1440
Qy 1441 TTCTTTAAACCCCAAGAGAGTGAAGCTGGAAGGCGATATCAGAGTCCCTCTTCCACA 1500
Db 1441 TTCTTTAAACCCCAAGAGAGTGAAGCTGGAAGGCGATATCAGAGTCCCTCTTCCACA 1500
Qy 1501 GCTGGCCGGGTGATTGGCAAGAGTGGCAAGCCGTGAACGAATCTGCAAACTTAACAGT 1560
Db 1501 GCTGGCCGGGTGATTGGCAAGAGTGGCAAGCCGTGAACGAATCTGCAAACTTAACAGT 1560
Qy 1561 GCGAAGTCAATGTCCTCTGTAACCAACCGCAAGTGAAGAAATGAGAAATGATGCTGGA 1620
Db 1561 GCGAAGTCAATGTCCTCTGTAACCAACCGCAAGTGAAGAAATGAGAAATGATGCTGGA 1620
Qy 1621 ATTATCGGCACTTCTTTGCTAGCACTGCAAGCGCAAGATCAGGAAAATTTGACAA 1680
Db 1621 ATTATCGGCACTTCTTTGCTAGCACTGCAAGCGCAAGATCAGGAAAATTTGACAA 1680
Qy 1681 CAGGTGAAGCAGCAGAGCAGAAATACCTCAAGGAGTGTGCTTCAAGGCGCAAGTGA 1740
Db 1681 CAGGTGAAGCAGCAGAGCAGAAATACCTCAAGGAGTGTGCTTCAAGGCGCAAGTGA 1740
Qy 1741 GGTCTCCCAAGGCAACAGCAAAACAAACGGAATGTAATGTAGCCCTTCCAAACCTGACAGAA 1800
Db 1741 GGTCTCCCAAGGCAACAGCAAAACAAACGGAATGTAATGTAGCCCTTCCAAACCTGACAGAA 1800
Qy 1801 TGAGAACCAACGACGACGACGATCGGAGCAAAACCAAGACCACTTGAAGATGAGAA 1860
Db 1801 TGAGAACCAACGACGACGACGATCGGAGCAAAACCAAGACCACTTGAAGATGAGAA 1860
Qy 1861 GTCTGCGAGGCGGCGGCGGAGGCTGTGCGAGGCTTCAAGGCGGCGGAGGAGG 1920
Db 1861 GTCTGCGAGGCGGCGGCGGAGGCTGTGCGAGGCTTCAAGGCGGCGGAGGAGG 1920
Qy 1921 GCGGGGAAGGTGACCGAGGTTTGCAGAAACCAAGGCGGCGGCTCCCGCCCGAGGG 1980
Db 1921 GCGGGGAAGGTGACCGAGGTTTGCAGAAACCAAGGCGGCGGCTCCCGCCCGAGGG 1980
Qy 1981 TTCTGAGGCTTCAAGCATCACTTCCATCCATCCATCCATCCATCCATCCATCCATCC 2040
Db 1981 TTCTGAGGCTTCAAGCATCACTTCCATCCATCCATCCATCCATCCATCCATCCATCC 2040
Qy 2041 CGCTATCCCTTTTATGTTGAATTAACATTAAGTGAACGTTTCAAGGCGGAGGAGT 2100
Db 2041 CGCTATCCCTTTTATGTTGAATTAACATTAAGTGAACGTTTCAAGGCGGAGGAGT 2100
Qy 2101 ACCCTTTTCTGTGGAAATGCTCTGTACATGTGTGTAATTAAGGAGGAGATG 2160
Db 2101 ACCCTTTTCTGTGGAAATGCTCTGTACATGTGTGTAATTAAGGAGGAGATG 2160
Qy 2161 TTAAGATATGTGCTGTGGTTTACACAGGCTGCTGCAAGCGGTAAATATTTTGAAT 2220
Db 2161 TTAAGATATGTGCTGTGGTTTACACAGGCTGCTGCAAGCGGTAAATATTTTGAAT 2220
Qy 2221 AATATATCAATTAATCACTAATCTCAATTTTAAATTTTAAATTTTAAATTTTAAAT 2280
Db 2221 AATATATCAATTAATCACTAATCTCAATTTTAAATTTTAAATTTTAAATTTTAAAT 2280
Qy 2281 TTAAGAGAAAGAGGCTTTTCTAGACTTAAAGAAATTAAGTCTTTGGAGGCTCAAG 2340
Db 2281 TTAAGAGAAAGAGGCTTTTCTAGACTTAAAGAAATTAAGTCTTTGGAGGCTCAAG 2340
Qy 2341 TGTAGAGAGAGCTTTGAGGCAACCGGCAAAAATTCACAGAGGAAATCTGTGGA 2400
Db 2341 TGTAGAGAGAGCTTTGAGGCAACCGGCAAAAATTCACAGAGGAAATCTGTGGA 2400
Qy 2401 AGGACACTCAGGAGTTCGTGATCACTGTGTATGTCAACAGAGGAAATCGTCTCT 2460
Db 2401 AGGACACTCAGGAGTTCGTGATCACTGTGTATGTCAACAGAGGAAATCGTCTCT 2460
Qy 2461 TGAAGAGAAACTGTGCACTCCATGCTGTAGCTCAATACCAATTTCTTTTGC 2520
Db 2461 TGAAGAGAAACTGTGCACTCCATGCTGTAGCTCAATACCAATTTCTTTTGC 2520

QY 901 AAAATCTTGGCAACAATGGCTTGGTTGGAAAGACTGATTTGGAAAAAGAGCAAAATTTG 960
Db 901 AAAATCTTGGCAACAATGGCTTGGTTGGAAAGACTGATTTGGAAAAAGAGCAAAATTTG 960
QY 961 AAGAAAAATTTGAACATGAACAGAGGACCAAGATTAACAATCTCATCTTTGCAAGGATTTGAGC 1020
Db 961 AAGAAAAATTTGAACATGAACAGAGGACCAAGATTAACAATCTCATCTTTGCAAGGATTTGAGC 1020
QY 1021 AATATCAACCCGGAAAAAGAACCTCATCTGTGAAGGGCAAGTTGAGAGCTGTGCGAGTGT 1080
Db 1021 AATATCAACCCGGAAAAAGAACCTCATCTGTGAAGGGCAAGTTGAGAGCTGTGCGAGTGT 1080
QY 1081 GAGATAGAGATTATGAAGAGAGCTGGGTGAGGCTTTGAAATGATGTGCTGTGAAC 1140
Db 1081 GAGATAGAGATTATGAAGAGAGCTGGGTGAGGCTTTGAAATGATGTGCTGTGAAC 1140
QY 1141 ACCCACTCCGAGTACTTCTCCAGCCGTGTAACCCCATCAACAGTTTGCCCGTTCGCGAT 1200
Db 1141 ACCCACTCCGAGTACTTCTCCAGCCGTGTAACCCCATCAACAGTTTGCCCGTTCGCGAT 1200
QY 1201 CATCACTCTTATCCAGAGCAGAGATTTGAAATCTTCTCATCCCAACCCAGGCTGTGGGC 1260
Db 1201 CATCACTCTTATCCAGAGCAGAGATTTGAAATCTTCTCATCCCAACCCAGGCTGTGGGC 1260
QY 1261 GCCATCATCGGGAAGAGGGGGCAACATCAACAGCTGGCGAGATTGCGCGAGCTCT 1320
Db 1261 GCCATCATCGGGAAGAGGGGGCAACATCAACAGCTGGCGAGATTGCGCGAGCTCT 1320
QY 1321 ATCAAGATTGCCCTGCGAGAGGCCCAAGCGTCAAGGAAAGATGTGTCATCATCCGGG 1380
Db 1321 ATCAAGATTGCCCTGCGAGAGGCCCAAGCGTCAAGGAAAGATGTGTCATCATCCGGG 1380
QY 1381 CCAACCGGAAGCCCATGTTCAAGGCCCAAGGACGATTTTGAGAACTGAAAGGGAAAAAC 1440
Db 1381 CCAACCGGAAGCCCATGTTCAAGGCCCAAGGACGATTTTGAGAACTGAAAGGGAAAAAC 1440
QY 1441 TTCTTTAACCCCAAGAAAGATGAAGTGAAGCTGAAAGCCGATTCAGATGTGCCCTTTCACA 1500
Db 1441 TTCTTTAACCCCAAGAAAGATGAAGTGAAGCTGAAAGCCGATTCAGATGTGCCCTTTCACA 1500
QY 1501 GCTGGCCGGGTGATTGGCAAAAGGTGGCAAGCCGTGAACGAACTGCAAACTTTAACCAAGT 1560
Db 1501 GCTGGCCGGGTGATTGGCAAAAGGTGGCAAGCCGTGAACGAACTGCAAACTTTAACCAAGT 1560
QY 1561 GCAGAAAGTCACTGTGCTGTGACCAAAAGCCAGATGAAGATGAAGAGTGTGTCAG 1620
Db 1561 GCAGAAAGTCACTGTGCTGTGACCAAAAGCCAGATGAAGATGAAGAGTGTGTCAG 1620
QY 1621 AATTATGGGGCACTTCTTGCTAGCCAGACTGACAGAGCGCAAGATCAGGGAAATTTGACAA 1680
Db 1621 AATTATGGGGCACTTCTTGCTAGCCAGACTGACAGAGCGCAAGATCAGGGAAATTTGACAA 1680
QY 1681 CAGGTGAAGCAGAGAGCAAGAAATACCTCAAGGGAGTGCCTTCAACAGCTGACAGAA 1740
Db 1681 CAGGTGAAGCAGAGAGCAAGAAATACCTCAAGGGAGTGCCTTCAACAGCTGACAGAA 1740
QY 1741 GGTCTCCCAAGGCAACAGCAAAAACAGGTGAATGTAAGCTTCAACAGCTGACAGAA 1800
Db 1741 GGTCTCCCAAGGCAACAGCAAAAACAGGTGAATGTAAGCTTCAACAGCTGACAGAA 1800
QY 1801 TGAGACCAAAACGAGCCAGATCGGAGCAAAACCAAGAACCATCTGAGGATGAGAA 1860
Db 1801 TGAGACCAAAACGAGCCAGATCGGAGCAAAACCAAGAACCATCTGAGGATGAGAA 1860
QY 1861 GTCTGCGAGAGCGGCAAGGAACTTGTGCGAGGCTTGAAGAACCCAGAGGCGCAAGGCG 1920
Db 1861 GTCTGCGAGAGCGGCAAGGAACTTGTGCGAGGCTTGAAGAACCCAGAGGCGCAAGGCG 1920
QY 1921 GCGGAGGAAGGTCAAGCCAGGTTTGCAAGAACCAAGAGCCCGGCTCCGCGCCCAAGGCG 1980
Db 1921 GCGGAGGAAGGTCAAGCCAGGTTTGCAAGAACCAAGAGCCCGGCTCCGCGCCCAAGGCG 1980
QY 1981 TTCTGAGAGGTTGAGCATCACTTCAACCATCCAGATCTCTGGAATCTCCACGA 2040

Db 1981 TTCTGAGAGGTTGAGCATCACTTCAACCATCCAGATCTCTGGAATCTCCACGA 2040
QY 2041 CGCTATCCCTTTTGTGTAACATATAGGTGAAGGTGTCAAGACCAAGCAAAATGCAAC 2100
Db 2041 CGCTATCCCTTTTGTGTAACATATAGGTGAAGGTGTCAAGACCAAGCAAAATGCAAC 2100
QY 2101 ACCCTTTTCTGTGGCAAAATGCTCTGTGACATGTGTGTACATATTGAAGAGGAAGT 2160
Db 2101 ACCCTTTTCTGTGGCAAAATGCTCTGTGACATGTGTGTACATATTGAAGAGGAAGT 2160
QY 2161 TTAAGATATGTGGCTGTGGGTTACACAGGGTGTCTGACAGGGTAATATTGGAAT 2220
Db 2161 TTAAGATATGTGGCTGTGGGTTACACAGGGTGTCTGACAGGGTAATATTGGAAT 2220
QY 2221 AATATATCAATTAATCAACTCAATTTTATCAATTTATTAATTTTCTT 2280
Db 2221 AATATATCAATTAATCAACTCAATTTTATCAATTTTATTAATTTTCTT 2280
QY 2281 TTAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGATAAAGTCTTGGAGGTCTCACGG 2340
Db 2281 TTAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGATAAAGTCTTGGAGGTCTCACGG 2340
QY 2341 TGTAGAGAGAGCTTTGAGGCTCACCCGCAAAATTTCAACCCGAGAGGAATCTCGCGGA 2400
Db 2341 TGTAGAGAGAGCTTTGAGGCTCACCCGCAAAATTTCAACCCGAGAGGAATCTCGCGGA 2400
QY 2401 AGGACATCAAGGCAAGTCTGATCACTGTATGTCAAGAGAAAGGATACCGTCTCT 2460
Db 2401 AGGACATCAAGGCAAGTCTGATCACTGTATGTCAAGAGAAAGGATACCGTCTCT 2460
QY 2461 TGAAGAGAAACTGTGTCACTCTGTATGCTGTCTAGCTCATCAACCACTTCTCTTGGC 2520
Db 2461 TGAAGAGAAACTGTGTCACTCTGTATGCTGTCTAGCTCATCAACCACTTCTCTTGGC 2520
QY 2521 TTCAACAGGTTTAACTGTTTGTGCAATGCTATATATTTCTGTCTCTCTGTT 2580
Db 2521 TTCAACAGGTTTAACTGTTTGTGCAATGCTATATATTTCTGTCTCTCTGTT 2580
QY 2581 TATCTCTCCCTCCCTCCCTCCCTCTCTCTCATCTCATCTTTGAAATTTCTCTCAT 2640
Db 2581 TATCTCTCCCTCCCTCCCTCCCTCTCTCTCATCTCATCTTTGAAATTTCTCTCAT 2640
QY 2641 CCCTCATCTCAATCCGATATCAAGCAACCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2700
Db 2641 CCCTCATCTCAATCCGATATCAAGCAACCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2700
QY 2701 GTATCATCAACCAAAAGGAACAAAGGAAACAAAGGAAACCAAGGCTCACTTCACTT 2760
Db 2701 GTATCATCAACCAAAAGGAACAAAGGAAACAAAGGAAACCAAGGCTCACTTCACTT 2760
QY 2761 GGTTCATCAAAAGAAAGAGTCAATGTATCTGTCTAGCGTTTTGAAGAGGAAACA 2820
Db 2761 GGTTCATCAAAAGAAAGAGTCAATGTATCTGTCTAGCGTTTTGAAGAGGAAACA 2820
QY 2821 GGAACCCACCAACCAACCAATCAACCAACCAAAAGAAAAATTTCAATGAAGAAATGT 2880
Db 2821 GGAACCCACCAACCAACCAATCAACCAACCAAAAGAAAAATTTCAATGAAGAAATGT 2880
QY 2881 AATTGTCTTTTGTGCAATTTTGTGTATGCAATCAATTTTCAAGAAATGATTTCTTTC 2940
Db 2881 AATTGTCTTTTGTGCAATTTTGTGTATGCAATCAATTTTCAAGAAATGATTTCTTTC 2940
QY 2941 TTTAAAAAAAATGTGAGAAAGTGAATTTTCAAGGTTGTGTCAGGCGGCTT 3000
Db 2941 TTTAAAAAAAATGTGAGAAAGTGAATTTTCAAGGTTGTGTCAGGCGGCTT 3000
QY 3001 AAAATTCACAGATTTTAAACGAGAAAAACACAGAGAAAGTCTCAGGTCTTTTAA 3060
Db 3001 AAAATTCACAGATTTTAAACGAGAAAAACACAGAGAAAGTCTCAGGTCTTTTAA 3060
QY 3061 CTTGAGACTTGTCTTGTGTTTCTTGAAGATTTTGAAGCTGATAGTGAAGCAT 3120

1501 ATCAACGGCCACCGGAAGCCAGTTCAAGGCCAGGAGCGATCTTTGGGAACGTGAAA 1560
1432 GAGGAAAACTTTTAACTCCCAAGAGAAGTGAAGCTGGAAAGCATATCAAGTGGCC 1491
1561 GAGGAAAACTTTTAACTCCCAAGAGAAGTGAAGCTGGAAAGCATATCAAGTGGCC 1620
1492 TCTTCCACAGCTGGCGGGGTGATGGCAAAAGGTGGCAAGCGTGAACGAACTGGCAAG 1551
1621 TCTTCCACAGCTGGCGGGGTGATGGCAAAAGGTGGCAAGCGTGAACGAACTGGCAAG 1680
1552 TTAAACAGTGCAGAAATCATCTGCTCTGTGAACAAACGCGAGATGAAAATGAGAAATG 1611
1681 TTAAACAGTGCAGAAATCATCTGCTCTGTGAACAAACGCGAGATGAAAATGAGAAATG 1740
1612 ATCTGCAAGATTAATGGGCACTTTCTTGTACCAAGTCTGGCAAGCGAAAGTCAAGGGA 1671
1741 ATCTGCAAGATTAATGGGCACTTTCTTGTACCAAGTCTGGCAAGCGAAAGTCAAGGGA 1800
1672 ATTTTAAACAGAGTGAAGAGAGAGAGAGAGAGAAATACCTCAAGGAGTGGCTCAAGCGC 1731
1801 ATTTTAAACAGAGTGAAGAGAGAGAGAGAGAGAAATACCTCAAGGAGTGGCTCAAGCGC 1860
1732 AGCAAGTGAAGTCTCCCAAGAGAGAGAGAGAGAGAGAAATCAACGATGAATGAGCTTCCAAAC 1791
1861 AGCAAGTGAAGTCTCCCAAGAGAGAGAGAGAGAGAGAAATCAACGATGAATGAGCTTCCAAAC 1920
1792 CTGACAGATGAAGTGAAG 1851
1921 CTGACAGATGAAGTGAAG 1980
1852 GAATGAGAGTCTGGGAG 1911
1981 GAATGAGAGTCTGGGAG 2040
1912 CGAG 1971
2041 CGAG 2100
1972 CCCCAGAGGCTTCTGCAAGGCTTCAAGCCATCACTTCAACATCCATCGAGATCTCTCTGAA 2031
2101 CCCCAGAGGCTTCTGCAAGGCTTCAAGCCATCACTTCAACATCCATCGAGATCTCTCTGAA 2160
2032 CCCCAGAGGCTTCTGCAAGGCTTCAAGCCATCACTTCAACATCCATCGAGATCTCTCTGAA 2091
2161 CCCCAGAGGCTTCTGCAAGGCTTCAAGCCATCACTTCAACATCCATCGAGATCTCTCTGAA 2220
2092 AAAATGCACACCTTTTCTGTGGCAAAATGCTCTGTACATGTGTACATATTGAAA 2151
2221 AAAATGCACACCTTTTCTGTGGCAAAATGCTCTGTACATGTGTACATATTGAAA 2280
2152 GGGAGAGTGTAAAGATATGTGGCTGTGGCTGTACACAGGGGTGCTGAGGGTAAATATAT 2211
2281 GGGAGAGTGTAAAGATATGTGGCTGTGGCTGTACACAGGGGTGCTGAGGGTAAATATAT 2340
2212 TTTAGAAATATATATCAATTAATCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 2271
2341 TTTAGAAATATATATCAATTAATCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 2400
2401 TTTTCTTTTAAAGAGAAAGAGAGGCTTTTCTAGACTTTTAAAGAAATTAAGCTTTTGGAG 2460
2332 GTCTCAGAGTGTAG 2391
2461 GTCTCAGAGTGTAG 2520
2392 CTGCGTGAAG 2451
2521 CTGCGTGAAG 2580
2452 CCGTCTCTTGAAG 2511
2581 CCGTCTCTTGAAG 2640

2512 TCTCTTGTCTCAAGAGTTTAACTGGTTTTTTCATATGCTATATATATCTCTGTCT 2571
2641 TCTCTTGTCTCAAGAGTTTAACTGGTTTTTTCATATGCTATATATATCTCTGTCT 2700
2572 CTCTCTGTATATCTCTCCCTCCCTCCCTCCCTCCCTCTTCTCTCATCTCATCTTTTGA 2631
2701 CTCTCTGTATATCTCTCCCTCCCTCCCTCCCTCCCTCTTCTCTCATCTCATCTTTTGA 2760
2632 TTTTCTCATCTCCATCATCTCAATCCGATCTACAGCAACCCCTCCCTCCCTCCCTCAAGCA 2691
2761 TTTTCTCATCTCCATCATCTCAATCCGATCTACAGCAACCCCTCCCTCCCTCCCTCAAGCA 2820
2692 GTGCTGTAGATTCATCATCAAG 2751
2821 GTGCTGTAGATTCATCATCAAG 2880
2752 CTTCACATCTGTACTCAAAAG 2811
2881 CTTCACATCTGTACTCAAAAG 2940
2812 AGGAAACAG 2871
2941 AGGAAACAG 3000
2872 AAAGATGATTTTGTCTTTTGTGATTTTGTGATTAAGCCATCATATTCAGCAAAATG 2931
3001 AAAGATGATTTTGTCTTTTGTGATTTTGTGATTAAGCCATCATATTCAGCAAAATG 3060
2932 ATTCCTTTCTTTAAAAAATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2991
3061 ATTCCTTTCTTTAAAAAATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3120
2992 CAGGAGGTTAAATTCACAGATTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3051
3121 CAGGAGGTTAAATTCACAGATTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180
3052 GTGTTTTTACCTCAGACCTTGTCTTGTGTTTCCCTTGAAGATTTTGAAGTGAATG 3111
3181 GTGTTTTTACCTCAGACCTTGTCTTGTGTTTCCCTTGAAGATTTTGAAGTGAATG 3240
3112 TTGAGACATTTTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3171
3241 TTGAGACATTTTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3300
3172 GCCAGCTGTGAAG 3231
3301 GCCAGCTGTGAAG 3360
3232 AGCCAAAGACAAATATGCTCTTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3283
3361 AGCCAAAGACAAATATGCTCTTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3412

RESULT 4
US-09-899-651-6
; Sequence 6, Application US/0989651
; Patent No. 6576756
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gare, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Kuntz, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
; TITLE OF INVENTION: Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/899,651
; CURRENT FILING DATE: 2001-07-06
; PRIORITY APPLICATION NUMBER: US/09/061,709

Db 241 CGGAACCTTCAGATCGAATATCCCGCTCATTTACAGTGGAGGTGTGATGTTTA 300
 QY 373 TTGGCTCAATATGAGACAGTGGAGATGTGAAACAAGTCAACAGACAGAAACCGCC 432
 Db 301 CTAGTCCAGTATGAGTGTGAGAGCTGTGACAGTGAACAATGACCTCGGAACTGCA 360
 QY 433 GTTGTCAACGTCAATATGCAACAAGAGAGAGAAATATGCAATGAGAGATTAAGC 492
 Db 361 GTTGTAAATGTAACTATTCAGTAAAGCAAGCTAAGCAAGCACTAGCAACTGAT 420
 QY 433 GGGCATCGTGTGAGAACTACTCTCTTCAAGATTTCTAATCCCGATGAAAGGTGAGC 552
 Db 421 GGATTTCAAGTAAAGAAATTTCACTTGAAGTAAAGTAACTCTGATGAAACGGCCGCC 480
 QY 553 TCCCTTGGCCCCCTCAGCGAGCCAGAGTGGGGACCACTCTCCCGGAGCAAGGC--- 609
 Db 481 CAGCAAAACCTTTGACAGAGCCCGAGGTGCGCGGGGCTTGGGCAAGAGGGGCTCTCA 540
 QY 610 -----CAGGCCCTGGGGGACCTTCTCAGGCGACAGATTTGATTTCCCGTGGGATC 663
 Db 541 AGGAGGGGTCTCCAGAGATCCGTATCCAGAGAGAAACATGTGATTTGCTCTGGCCTG 600
 QY 664 CTGCTCCCAACCCAGTTTGTGTGTCATCGGAAAGAGGGCTTGACCAATAAAGAC 723
 Db 601 CTGATTCACCCCAATTTGTGAGCCATCATAGGAAAGAGGTGCGACCATTCGAGAC 660
 QY 724 ATCACTAAGCAGACCCAGTCCCGGGTAGATTCATAGAAAGAGAACTCTGAGCTGCA 783
 Db 661 ATCCCAAAAGAGCCAGCTTAAATCATGTATCCAGCTAAAGAAATGCGGGGCTGCT 720
 QY 784 GAGAGGCTGTCACTTCATCCAGCCAGAGGGGACCTTGAAGATGCGGATAT 843
 Db 721 GAGAAAGTCAATATCTCTCTCTCACTCTGAGAGCACCTTGGGCTTTGAACTGAT 780
 QY 844 CTGTAATCATGCAAGAAAGGAGATGAGCAACAACCTGCGGAGATTCCTCTGAAA 903
 Db 781 CTGAGATTTATGCTAAGAAAGCTCAAGATTAATAATTCACAGAGATCCCTTGAG 840
 QY 904 ATCTGGCACAATGCTGTGTGAGAACTGATTTGAAAGAGAGCAAAATTTGAG 963
 Db 841 ATTTTACTCATATTAATCTTTGTGAGCTTTATTTGTAAAGAGAGAAATCTTAAA 900
 QY 964 AAAATGACATGAAAGAGGAGCCAGATTAACAATCTCATTTGAGAGATTTGAGATA 1023
 Db 901 AAAATGAGAGAGCAAGACATTAATAATCAAGATTTCTCATTTGAGAGATTTGAG 960
 QY 1024 TACAAACCGGAAAGACATCACTGTGAGAGGCAAGTTGAGGCTGTGCAAGTGTAG 1083
 Db 961 TATATATCAGAGAGCACTATTAAGGCAATGTGAGACATGTGCCAAAGCTGAG 1020
 QY 1084 ATGAGATTTATGAAAGAGCTGCGTGAAGCCTTTGAAATGATATGCTGCTGTTAACC 1143
 Db 1021 GAGAGATTCATGAAAGAAATCAGGAGTCTTATGAAATATATTTGCTTTATGAACTT 1080
 QY 1144 CACTCCGAT-----ACTTCCCA 1162
 Db 1081 CAGACATTTAATTCCTGATTAATCTGAAGCCTTGGGTCTGTTCCACCACTTCA 1140
 QY 1163 GCTGTAAACCCCATCAACAGTTTGGCCGTTCCCGATCACTC-----ACTTCCCA 1208
 Db 1141 GGGATGCACTCTCCACTCAGAGGCCCCCTTCAAGCCATGACTCTCCCTACCGCAGTT 1200
 QY 1209 --TTATCCAGAGAGATTTGATTTCTTATCTTATCCCAACCAAGGTGTGGGCGCATC 1266
 Db 1201 GAGCAATCGAAACGAGAGCTGTTCACTGTTTATCCAGCTTATCAAGCTGATGATC 1260
 QY 1267 ATCGGAAAGAGGGGACACATCAACAGCTGCGGAGATTTGCGCGGAGCTTATCAAG 1326
 Db 1261 ATCGGAAAGAGGGGACACATCAAGCACTTCTGCTTGGAGCTTAAATTTAG 1320
 QY 1327 ATTGCCCTTGGGAGGCCAGAGCTCAGGAAAGATGATCATCACCGGAGCAACG 1386

Db 1321 ATTGCTCAGGAGAGACACAGATGCTAAAGTGAAGATGTGATTTACTGTGACACCA 1380
 QY 1387 GAGCCCACTTCAAGGCCAGGAGCGATCTTTGGGAAACTGAAAGAGAAACTCTTT 1446
 Db 1381 GAGGCTCAGTTCAAGGCTCAAGGAAAGATTTATGGAATAATTAAGAAAGAACTTTGT 1440
 QY 1447 AACCCAAAGAAAGTAACTGGAAGCGATATCAAGTGGCTCTTCCAGAGCTGGC 1506
 Db 1441 AGCTTAAAGAGGTAAACTTGAAGCTCATATCAAGAGGCCATTCCTTTGCTGTGC 1500
 QY 1507 CGGATATTTGCAAAAGGTGCAAGACCGTGAACGAATCGCAAACTTAAACGATCAGAA 1566
 Db 1501 AGAGTTATGGAAGAAAGAGGCAAAACGATGATTAATCTGAAATTTGTCAAGTCAGAA 1560
 QY 1567 GTCATGTGCTCTGTACCAACCGCAGATGAAATGAGAAAGTATGTCAGAAATTAAC 1626
 Db 1561 GTTGTGTCTCTGTGACCAACAGCTGATGAGATGACCAAGTGTGTGAAATTAAC 1620
 QY 1627 GGGGACTTTTGTGACCAACAGCTGCAACGCGCAAGATCAGGAAATTTGTCACACAGTG 1686
 Db 1621 GGTCACTTATGCTTGTGCAAGTGTGCCAGGAAATTAAGGAAATTTGATCACTCAGGTA 1680
 QY 1687 AAGCAGAGAGCAGAAATACCTCAGGAGTGTGCTCAG 1726
 Db 1681 AAGCAGACCAACAAGAGGCTTCAAGTGAACAC 1720
 RESULT 6
 US-09-542-615A-347
 ; Sequence 347, Application US/09542615A
 ; Patent No. 6518256
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Tongtong
 ; APPLICANT: Fan, Liqun
 ; APPLICANT: Kalos, Michael D.
 ; APPLICANT: Bangur, Chaitanya S.
 ; APPLICANT: Hoeken, Nancy A.
 ; APPLICANT: Fanger, Gary R.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
 ; FILE REFERENCE: AND DIAGNOSIS OF LUNG CANCER
 ; CURRENT APPLICATION NUMBER: US/09/542,615A
 ; CURRENT FILING DATE: 2000-04-14
 ; NUMBER OF SEQ ID NOS: 350
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 347
 ; LENGTH: 1740
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-542-615A-347
 Query Match 21.2%; Score 697.2; DB 4; Length 1740;
 Best Local Similarity 64.9%; Pred. No. 4.8e-185;
 Matches 1116; Conservative 0; Mismatches 538; Indels 66; Gaps 3;
 QY 73 ATGACAAAGCTTTACATCGGAAACCTGAGCCCGCGTCAACCGGACGACTCCGCGAG 132
 Db 1 ATGAAACAACTGATATCGGAAACCTCAGGAGAGCGCCCGCTCGACCTTAGAAAGT 60
 QY 133 CTCTTGGGAGACAGAAAGCTGCCCTGGGGGAGCAGGTCTGCTGAAGTCCGGTACGCC 192
 Db 61 ATCTTCAAGAGACCCAAAGATCCGGTGTGGAGACCTCTCTGTGAGAGCTGTACGCG 120
 QY 193 TTGCTGAGACTACCCGACCAAGAACTGGGCAATCCGCGCATGAGACCTCTGGGTA 252
 Db 121 TTGCTGAGAGTCCCGGACGAGGCTGGGCTCAAGGCAATGAGGCTTTTACGTA 180
 QY 253 GTGAGATTTGATGGAATAATCATGAAAGTTGATTTCACTCTTAAAGGCTTAAGAGC 312
 Db 181 ATGAACTCAGCGGAAACCATTAAGAGTTGACACTCGCTCCCAAAAGGCAAGGAT 240
 QY 313 AGGAAATTCAGATTTGAAACATCCCTCTCAGCTGAGGAGGTGTTGATGACCT 372

```

Db      241 CGGAATCTTGAGATACGAAATATCCCGCTCATTTACAGTGGAGGTGTGATAGTTTA 300
Qy      373 TTGGCTCAATATGGGACAGTGGAGATGTGAAACAAGTCAACAAGACACAGAAACCGCC 432
Db      301 CTAGTCCAGTATGGAGTGTGGAGAGCTGTGAGCAAGTGAACACTGACTCGGAAACTGCA 360
Qy      433 GTTGTCAACGTCACTATGCAACAGAGAAAGAAATATAGCCATGGAGAGCTTAAGC 492
Db      361 GTTGTAAATGTAACTTATTCAGTAGAGCCAGCTTAACAGACACTAGACAACTGAT 420
Qy      493 GGGCATCAGTTTGAAGAACTACTCTTCAAGATTTCTCAATCCCGATGAAAGGTGAGC 552
Db      421 GGATTTCACTTGAAGATTTCACTTGAAGTACCTTATTCCTGATGAAACCGCCGCC 480
Qy      553 TCCCTTGGCCCTTCAAGCAGCCAGCGTGGGAGCACTCTTCCCGGAGCAAGGC--- 609
Db      481 CAGCAAAACCCCTTGCAGCAGCCCGAGGTGCGCCGAGGCTTGGGCAAGGAGGCTTCA 540
Qy      610 -----CAGCCCTTGGGGGCACTTCTAGGCGCAGACAGATTGTTCCCGCTGGGATC 663
Db      541 AGGCAAGGGGTTCAGAGATCGTATTCAGAGAAACCATGTGATTTGCTTGGCGCTG 600
Qy      664 CTGCTCCCACTTGTGTTGTGTGTCATTCGAAAGAGGCTTGAACATGAAAGAAC 723
Db      601 CTGCTTCCCACTTGTGTTGTGTGATTCAGAGAAAGAGGTGCGCACATTCGGAAC 660
Qy      724 ATCACTAAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAGAAACCTTGGAGCTGCA 783
Db      661 ATCACTAAGCAGACCCAGTCTTAAATTCATATGTCACCGTAAAGAAATGCGAGGCTGT 720
Qy      784 GAGAGCCGTGACCATGCAATGCAACCCGAGGAGCTTGAAGCATGCGCATGAT 843
Db      721 GAGAGGTGATTAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 780
Qy      844 CTTGAATCATGACAGAAAGGAGAGTGAACCAATGACCCGAGAGATTCCTCTGAAA 903
Db      781 CTGGAGATTAATGATTAAGAGAGCTCAAGATTAATTAATTAATTAATTAATTAATTA 840
Qy      904 ATCTTGGCAACAATGGCTTGTGTGAAACATGATTTGAAAGAAAGGAGAAATTTGAAG 963
Db      841 ATTTTATGCTCATATATTAATCTTGTGTGAGCTTATTTGTGTAAGAAAGAAATCTTAAA 900
Qy      964 AAAATGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1023
Db      901 AAAATGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 960
Qy      1024 TACAACTCCGAGAAAGAACATCACTGTGAAGGAGCAAGTGAAGGCTGTGCAAGTCTGAG 1083
Db      961 TATTAATCCAGAAAGCACTATTAAGTTAAAGCAATGTGAGACATGTGCCAAAGCTGAG 1020
Qy      1084 ATAGAGATTAATGAAGAGCTGCTGAGGCTTTGAAATTAATGATGCTGCTTAAACACC 1143
Db      1021 GAGGAGATATGAAGAAATCAGGAGTCTTATGAAGAAATGATGCTTAAAGAACTTT 1080
Qy      1144 CACTCCGGAAT-----ACTTCTCCA 1162
Db      1081 CAGAGCATTTAATTTCTGATTAATTTGAAGCGCTTGGGTCTGTTCCCAACCACTTCA 1140
Qy      1163 GCTCTAATCCCACTCAACCACTTTGGCCGCTTCCCGCATCACTC----- 1208
Db      1141 GGGATGCACTCTCCCACTCAAGGAGCCCTTCAAGCCATGACTCTCTCCCTTAACCGCAGTTT 1200
Qy      1209 --TTATTCAGAGAGAGATTTGTAATCTCTCATCCCAACCGAGGTGTGGGCGCATC 1266
Db      1201 GAGCAATCAAGAAACGAGACCTTCTATCTTTATCCAGCTTATATAGTCGGTGCATC 1260
Qy      1267 ATGGGAGAAAGAGGAGAGACATCAAAACAGCTGGCGAGATTTCCCGGAGCTTATCAAG 1326
Db      1261 ATGGGAGAGAGAGGAGAGACATCAAGACACTTCTTCTGCTTGTGAGGCTTCAATTAAG 1320
Qy      1327 ATTTGCCCTTGGAGAGGCGCAGAGCTGACGAGAAAGATGTCTATCAACCGGCGCACCG 1386
Db      1321 ATTTGCTCAAGGAGAGCAACAGATGTAAAGTGAATGTGATTAATCACTGAGCACCA 1380

```

```

Qy      1387 GAAGCCAGTTCAAGGCCCAAGGACGATCTTTGGGAACTGAAAGAGGAAATCTTCTT 1446
Db      1381 GAGGCTCAGTTCAAGGCTCAAGGAAAGATTTATGGAATAATTAAGAAAGAAACTTGT 1440
Qy      1447 AACCCAAAGAAAGTGAAGCTGTGAAGCCATATCAAGTGCCTCTTCCAGCTGAGC 1506
Db      1441 AGCTTAAAGAAAGAGTGAAGCTTGAAGCTCATATCAAGTGCCTCTTGTGCTGCTG 1500
Qy      1507 CGGGTGAATGGCAAGGTGGCAAGACCGTGAACGACTGACAGAACTTAACAGTGCAGAA 1566
Db      1501 AGAGTTATGGAAAGAGGAGGAAACGATGATGAATGAACTTCAGAAATTTGTCAAGTCA 1560
Qy      1567 GTCAATGCTGCTGTGACCAACCGCAGATGAAATGAGAAAGTGAATGTCAAGAAATATC 1626
Db      1561 GTTGTGTCCCTCGTGCAGACACTGATGATGATGACAAAGTGTGTGCAAAATTAAT 1620
Qy      1627 GGGCATCTTCTTGTCTAGGCAAGTCAACAGGCAAGATCAGGGAATTTGTAACAGGTG 1686
Db      1621 GGTCACTTCTATGCTTGTGCAAGTGTGCCAGAGAAATTAATCAAGAAATTTCTGACTGAGTA 1680
Qy      1687 AAGCAGAGAGAGCAAGAAATACCTCAGGAGTGCCTGAC 1726
Db      1681 AAGCAGACCAACAAACAGAGCTTGCAAAGTGAACAC 1720

```

```

RESULT 7
US-09-606-421B-347
; Sequence 347, Application US/09606421B
; Patent No. 6531315
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Skelky, Yasir A.W.
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C9
; CURRENT APPLICATION NUMBER: US/09/606,421B
; CURRENT FILING DATE: 2000-06-28
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 347
; LENGTH: 1740
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-606-421B-347

```

```

Query Match      21.2%; Score 697.2; DB 4; Length 1740;
Best Local Similarity 64.9%; Pred. No. 4.8e-185;
Matches 1116; Conservative 0; Mismatches 538; Indels 66; Gaps 3;

Qy      73 ATGAACAGCTTATACATCGGAACTGAAGCCCGCGCTCAACGCGCAGCACTTCGGCAG 132
Db      1 ATGAACAAATCTGATATCGGAAACTCAAGCAAGAGCGCGCCCTTCGAGCACTGAAGAT 60
Qy      133 CTCTTTGGGAGAGAACTGCTCCCTGGCGGAGCAGTCTCTGTGAAGTCCGGCTACGCC 192
Db      61 ATCTTCAAGGACCCCAAGATCCCGGTGTGGAGCCCTTCTGTGGAAGACTGGCTACGCG 120
Qy      193 TTGGTGACTAACCCCGACGAACTGGCCATCCGCGCATTCAGACCTCTCGGTTAAA 252
Db      121 TTGTGACCTGCGCGAGCAGAGCTGGCCCTCAAGGCCATCAAGCGCTTTCAAGTAAA 180
Qy      253 GTGGAATTCATGGGAAATCATGAAGTTGATTACTCACTCTTAAAGCTTAAGAGC 312
Db      181 ATGAATCTGACCGGAAACCATTAAGAGTTGAAGACTCGGTCCCAAAAGCAAGAT 240

```


QY 313 AGAATAATTCAGATTGGAACATCCCTCCTCACTGAGTGGAGGTGTGGATGACTT 372
DB 241 CGAATACTCAGATAGGAATATCCGGCTCAATTACATGGAGAGTGTGATGATTTA 300
QY 373 TTGGCTCAATATGGACAGTGAAGATGTGAACAAGTCAACAGACACAGAAAACGCC 432
DB 301 CTAGTCCACTATAGGTGTGAGCTGTGAGCAAGTGAACATGACTCGAATCTGCA 360
QY 433 GTTGTCAAGTCACATATGCAACAAGAGAAGAAATATGCCATGGAAGAGTAAAGC 492
DB 361 GTTGTAAATGTAACTATTCCTAGTAAAGACAGACTGACAGCACTGACAACTGAT 420
QY 493 GGGCATCAAGTTGAGAACTACTCTTCAAGATTTCTCATCCTCCGATGAAGAGTGAAC 552
DB 421 GGATTTAGATTAGAAATTTCACTTGAAGAGTCTATATCTCGATGAAGACGCCGCC 480
QY 553 TCCCTTGGCCCCCTCAGAGAGCCAGAGTGGGAGCACTCTTCCGGAGAGCAAGC--- 609
DB 481 CAGCAAAACCCCTTGGCAGAGCCCGAGGTGCGCCGGGGCTTGGGCAAGAGGGCTCTCA 540
QY 610 -----CAGGCCCTGGGGGCACTTCTAGGCGCAGACAGATTGATTTCCCGCTGGATC 663
DB 541 AGGAGAGGGGTCTCAGAGATCCGTATCAGAGAAACCATGTGATTTGCTGTGGCCTG 600
QY 664 CTGTCTCCCAACCAAGTTTGTGTGATCCATCCGAAAGAGAGGCTTGACCATAAAGAC 723
DB 601 CTGTCTCCCAACCAAGTTTGTGTGATCCATCCGAAAGAGAGGCTTGACCATTCGAGAC 660
QY 724 ATCACTAAGCAGACCCAGTCCCGGGTGAATATCCATAGAAAAGAACTTGGAGCTGCA 783
DB 661 ATCACTAAGCAGACCCAGTCCCGGGTGAATATCCATAGAAAAGAACTTGGAGCTGCT 720
QY 784 GAGAAGCCTGTACCATCATGCAACCCAGAGGGAGATTGTGAAGCATGCGCATGATT 843
DB 721 GAGAAGCCTGTACCATCATGCAACCCAGAGGGAGATTGTGAAGCATGCGCATGATT 780
QY 844 CTTGAAATCATGAGAAAGAGAGAGATGAGACCAACTGAGCGAAGATTCCTGTGAA 903
DB 781 CTGAGATTTATGATAGAAAGAGAGATGAGATGAGATGAGATGAGATGAGATGAG 840
QY 904 ATCTGACACAAATGCTGCTGTGTGAAGATGAGATGAGATGAGATGAGATGAGATGAG 963
DB 841 ATTTTACTCATATATTAATTTGTGAGAGCTTATGATGATGATGATGATGATGATGAT 900
QY 964 AAAATGACATGAGAAAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 1023
DB 901 AAAATGACATGAGAAAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 960
QY 1024 TACAAACCGAGAAAGACATCATCTGTAGAGGCAAGTTGAGGCTGTGACAGCTGAG 1083
DB 961 TATATATCAGAGAGAGATTTAAGTAAAGGCAATGTTGAGATGAGATGAGATGAGATGAG 1020
QY 1084 ATGAGATTTATGAGAAAGCTGCTGAGGCTTGTGAATGATGCTGCTGTGAACACC 1143
DB 1021 GAGAGATGATGAGAAAGCTGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 1080
QY 1144 CACTCCGAGT-----ACTTCTCCA 1162
DB 1081 CAGAGCATTTAATCTCTGATTAATGTAAGCCTTGGGTCTGTTCACCAACCATTTCA 1140
QY 1163 GCGCTGACCCCATCAGATTTGCGCGTTCCCGCATCATCTC----- 1208
DB 1141 GGGATGCACTCCCACTCAGAGGCCCCCTTTCAGCCATGACTCTCCCTACCCGAGATT 1200
QY 1209 ---TTATCCAGAGAGAGATTTGAATCTCTTCATCCCAACCCAGAGCTGTGAGGAGCATC 1266
DB 1201 GAGCAATTCAGAAACGAGAGCTGTTCACTGTTATATCCAGACTCTATCAGTCCGTGACATC 1260
QY 1267 ATGCGAAGAGAGGAGAGACATCAACAGCTGAGAGATTTGCGCGAGCTCTATCAAG 1326
DB 1261 ATGCGAAGAGAGGAGAGACATCAACAGCTGAGAGATTTGCGCGAGCTCTATCAAG 1320
QY 1327 ATTGCCCCCTGCGAGAGGCCAGAGCTCAGCGAAAGAGATGTCATCAACCGGAGCAGCG 1386

DB 1321 ATTGCTCAGCGGAGACACAGATCTAAAGTGAAGATGTGATATCACTGAGACCA 1380
QY 1387 GAAGCCAGTTCAAGGCCAGGAGAGAGATCTTTGGGAAATGAAAGAGAAATCTCTTT 1446
DB 1381 GAGGCTCAGTTCAAGGCTCAGGAGAGAAATTTATGAGAAATTTAAAGAGAAATCTTTGTT 1440
QY 1447 AACCCAAAGAGAGAGTGAAGCTGAGAGCGATATCAGAGTCCCTCTTCCACAGCTGAGC 1506
DB 1441 AGTCCAAAGAGAGAGTGAAGCTGAGAGCGATATCAGAGTCCATCTTGTGCTGTGAGC 1500
QY 1507 CGGATGATTTGGAAAGGTGGCAAGCCGTGAACGAATCTGACAACTTAAACAGTCAAGAA 1566
DB 1501 AGATTAATGAGAAAG 1560
QY 1567 GTCATCTGCTCTGAGCAACCAAGCCAGATGAAATGAGAGAGAGAGAGAGAGAGAGAGAG 1626
DB 1561 GTTGTGTCTCTGAGCAACCAAGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 1627 GGGCACTTCTTGTCTAGCAG 1686
DB 1621 GGTCACTTCTATGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
QY 1687 AAG 1726
DB 1681 AAG 1720

RESULT 8
US-09-061-709-4
; Sequence 4, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tseng, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 4
; LENGTH: 4159
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-4

Query Match 21.2%; Score 695.6; DB 3; Length 4159;
Best Local Similarity 64.8%; Pred. No. 2,5e-184;
Matches 115; Conservative 0; Mismatches 539; Indels 66; Gaps 3;

QY 73 ATGAACAAGCTTATCATCGGAGAGCTGAGAGCCCGGCTACCGCGAGAGAGAGAGAGAG 132
DB 251 ATGAACAAGCTTATCATCGGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 310
QY 133 CTCTTTGGGAGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 192
DB 311 ATCTTCAAG 370
QY 193 TTCTGTGAG 252
DB 371 TTCTGTGAG 430
QY 253 GTGGAATTCATGAG 312
DB 431 ATGAAG 490

QY 313 AGGAAATTCAGATTGGAACATCCCTCTGACCTGAGTGGGAGTGTGGATGAGCTT 372
 DB 491 CGGAAATCTTGAATGATGACAAATATCCCGCTATTTCAGATGGGAGGCTGGATGATTTA 550
 QY 373 TTGGCTCAATGTGGAGCAGTGGAGAAATGTGGACAAGTCAACAGACACAGAAACCGCC 432
 DB 551 CTAGTCCAGTATGAGTGTGGAGAGCTGTGAGCAAGTGAACCTGACTCGGAAATCTCA 610
 QY 433 GTTGTCAACGTCAATATGCAACAAGAGAGAGCAAAATAGCCATGAGAAAGCTTAAGC 492
 DB 611 GTTGAATGATGACCTATTCAGTAAGGACCAAGCTACAGACAGCACTAGCAAACTGAT 670
 QY 493 GGGCATGAGTTTGAAGTACTCTCTTCAAGATTTCTCATCTCCGGATGAGAGGTGAGC 552
 DB 671 GGAATTCAGTGAAGAAATTTACCTTGAAGTAGCTTATTCCTGATGAATGGCCGCC 730
 QY 553 TCCCTTGGCCCCCTCAGCGAGCCAGCGTGGGAGCAACTCTTCCCGGAGCAAGGC--- 609
 DB 731 CAGCAAAACCCCTTGACAGAGCCCGAGGTGCGCCGGGGCTTGGGCAAGAGGGCTCTCA 790
 QY 610 -----CAGCCCTCGGGGGGCACTTCTAGAGCCAGACAGATTTTCCCGTGGGAGTC 663
 DB 791 AGGCAAGGGGTCTCCAGATCCGATCCAGCAAGAACCATGATTTGCTTGGCCCTG 850
 QY 664 CTGCTCCCAACCCAGTTTGTGTGTCATCATCGGAAAGAGGGCTTGACATAAGAAC 723
 DB 851 CTGCTCCCAACCCAGTTTGTGTGTCATCATCGGAAAGAGGGCTTGACATAAGAAC 910
 QY 724 ATCACTAAGCAGACCCAGTCCCGGGTGAATATCATAGAAAAGAGAACTCTGAGCTGCA 783
 DB 911 ATCAACCAACACACCCAGTCAAAATCATGATGTCACCCGTAAGAAATGCGGGGCTGT 970
 QY 784 GAGAAAGCTGTCAACATTCATGCAACCCAGAGGGAGCTTGGAAAGCAAGCCGANTGAT 843
 DB 971 GAGAAAGCTGTCAACATTCATGCAACCCAGAGGGAGCTTGGAAAGCAAGCTTAT 1030
 QY 844 CTGGAATATGATGAGAAAGAGGAGATGAGACCAACTGCGGAGAGATCTCTGAAA 903
 DB 1031 CTGGAATATGATGAGAAAGAGGAGATGAGACCAACTGCGGAGAGATCTCTGAAA 1090
 QY 904 ATCTTGGCAGACATGGCTTGTGTGGAAGCTGATTTGAAAAGAGCAAAATTTGAAG 963
 DB 1091 ATTTTAGCTCATTAATCTTTGTGTGAGCTTTATTTGTAAGAGAGAAATCTTAAA 1150
 QY 964 AAAATGGAACATGAACAGGAGCAAGATTAACATCTCATCTTGGAGATTTGACATA 1023
 DB 1151 AAAATGGAACATGAACAGGAGCAAGATTAACATCTCATCTTGGAGATTTGACATA 1210
 QY 1024 TACAAACCCGAAAGAACATCACTGTGAAGGGAGAGGCTGTGCGAGTGTGAG 1083
 DB 1211 TATATATCGAAGAGCACTATTATTAAGGCAATGTTGAGACATGTGCCAAAGCTGAG 1270
 QY 1084 ATAGAGATTATGAGAGAGCTGCTGAGGCTTGAAGATGATGCTGCTGTTAACACC 1143
 DB 1271 GAGAGATATGAGAAAGAAATCAGGAGGTCTTATGAAGATATGCTTATGAAATCTT 1330
 QY 1144 CACTCCGAGT-----ACTTCTCCA 1162
 DB 1331 CAGGACATTTAATTCCTGATTAATCTGAAGCTTGGGTCTGTTCCACCCACTTCA 1390
 QY 1163 GCGCTAGCCCCCATCAACAGTTTGGCCGTTCCCGCATCATCTC-----ACTTCTCCA 1208
 DB 1391 GGGATGCCACTCCCACTTCAAGGCCCCCTTGACGCATGACTCTCCCTTACCCGCACTT 1450
 QY 1209 --TTATTCAGAGCAGAGATTTGTAATCTCTTCACTCCCAACCAAGCTGTGGGCGCATC 1286
 DB 1451 GAGGAATCAGAAAGAGAGCTGTTCATCAATTAATCCAGTCTATCAAGTCGGGCGCATC 1510
 QY 1267 ATGGGAGAGAGGGGAGACATTAACAAGTGGCGAGATTTCCGCGAGGCTTATCAAG 1326
 DB 1511 ATCGGCAAGAGGGGCGACACATCAAGAGCTTCTGCTTGTGAGAGCTTCAATTAAAG 1570
 QY 1327 ATTGCCCCCTGGGAAGGCCAGACGTACAGGAAGAGATGTCATCATCAACGGGCGACCG 1386

DB 1571 ATTGCTCAGCGGAGAGCAGACAGATGCTTAAAGTGAAGTGTGATTTATCACTGACCA 1630
 QY 1387 GAAGCCAGATTCAAGAGCCAGGACGAGATCTTTGGGAACTGAAGAGAGAACTTCTT 1446
 DB 1631 GAGGCTCAATTCAGAGCTCAGGAGAAATTTATGAGAAATTTAAGAGAAACTTGT 1690
 QY 1447 AACCCAAAGAGAAAGTGAAGCTGAGAACCCGATTCAGAGTGTGCTCTTCCACAGCTGGC 1506
 DB 1691 AGTCTTAAAGAGAGGTGAATCTTGAAGCTATTCAGAGAGGCCATCTTTGTGCTGGC 1750
 QY 1507 CGGATGATTTGGCAAGGTGGCAACCGTGAACCACTGACGAACCTTAAACAGTGCAAG 1566
 DB 1751 AGAGTTATTTGAGAAAGAGAGGCAAAACGCTGATGAACTTCAAGATTTGTCAAGTGCA 1810
 QY 1567 GTCATGTCCTCGTGAACCAACCGCAGATGAAATGAGAAAGTGAATGTCAGAAATTATC 1626
 DB 1811 GTTGTGTCTCTCGTGAACCAACCGCAGATGAAATGAGAAAGTGAATGTCAGAAATTATC 1870
 QY 1627 GGGCACTTCTTGTGTAAGCACTGACAGAGGCAAGATCAAGGAATTTGTACACAGGTG 1686
 DB 1871 GGTCACTTCTATGCTTGTGCGAGGTGCGCAGAGAAATTCAGAAATTTCTGACTAGGTA 1930
 QY 1687 AAGCAGCAGAGCAGAAATACCTCAGGAGTGTGCTCAC 1726
 DB 1931 AAGCAGCAGCAGAAACAGAGAGGCTCTGCAAAAGTGAACAC 1970

RESULT 9

US-09-899-651-4
 ; Sequence 4, Application US/09899651
 ; Patent No. 6576756

; GENERAL INFORMATION:
 ; APPLICANT: Chen, Yao-Tseng
 ; APPLICANT: Gure, Ali
 ; APPLICANT: Tseng, Elisabeth
 ; APPLICANT: Stockert, Elisabeth
 ; APPLICANT: Jager, Elke
 ; APPLICANT: Knuth, Alexander
 ; APPLICANT: Old, Lloyd J.
 ; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
 ; TITLE OF INVENTION: Associated
 ; FILE REFERENCE: LUD 5538
 ; CURRENT APPLICATION NUMBER: US/09/899,651
 ; PRIOR FILING DATE: 2001-07-06
 ; PRIOR APPLICATION NUMBER: US/09/061,709
 ; PRIOR FILING DATE: 1998-04-17
 ; NUMBER OF SEQ ID NOS: 8
 ; SEQ ID NO 4
 ; LENGTH: 4159
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:

US-09-899-651-4
 Query Match 21.2%; Score 695.6; DB 4; Length 4159;
 Best Local Similarity 64.8%; Pred. No. 2,5e-184;
 Matches 115; Conservative 0; Mismatches 539; Indels 66; Gaps 3;

QY 73 ATGAACAAGCTTAAATCCGGAACCTGAGCCCGCGCTCAACCGCGACGACTCCGGCAG 132
 DB 251 ATGAACAAGCTTAAATCCGGAACCTGAGCAAGAGAGCGCCCGCTCGACCTAAGAAAGT 310
 QY 133 CTCTTGGGAGCAGAGAGCTGCCCTTGGCGGAGCAGGTCTCTGTGAAGTCCGGCTACGCC 192
 DB 311 ATCTTCAAGAGCGCAAGATTCGGGTGTGGGACCTTCTCTGTGAAGACTGGCTACGCG 370
 QY 193 TTGGTGAATACCCCGACAGAACTGGGCTATCGCGCCATTCAGAGACCTCTCGGGTAAA 252
 DB 371 TTGGTGAATACCCCGACAGAGCTGGGCTCTCAAGGCGCAATCGAGGCGCTTCAAGTAAA 430
 QY 253 GTGAATTTGATGGGAAATCATGAAAGTTGATTAATCACTCACTTAAAAAGCTAAGAGC 312

Db 431 ATGAACTGACGGGAAACCCATGAAAGTTGACACTCGGTCCCAAAAGGCAAGAGATT 490
Qy 313 AGGAAATTCAGATTGAAACATCCCTCCATCCTCAGTGGAGAGGTGTTGATGACATT 372
Db 491 CGGAAATTCAGATTGAAATATCCGCTCATTTACAGTGGAGAGGTGTTGATGATT 550
Qy 373 TTGGCTCAATATGAGACAGTGGAGATGTGAACAAGTCAACACAGACACAGAAACGGC 432
Db 551 CTAGTCCAGTATGAGAGTGTGGAGAGCTGTGACAGATGAACATGACTCGAAATGTGA 610
Qy 433 GTTTCACAGTCACTATGCAACAGAGAAAGCAAAATAGCCATGAGAGAGTAAAGC 492
Db 611 GTTGAATATGTAACCTATTCAGTAAAGCACCAAGTGAACAGCACTAGACAACTGAT 670
Qy 493 GGGCATCAGTTGAGACACTCTCTCAAGATTCTTACATCCGAGTAAAGGTGAC 552
Db 671 GGATTCAGTTAGAAATTTACCTTGAAAGTAGGCTTATCTCGATGAATATGACCGCC 730
Qy 553 TCCCTTGGCCCCCTCAGGAGCCGAGCGTGGGAGCACTCTTCCCGGAGCAAGGC--- 609
Db 721 CAGCAAAACCCCTTGACAGAGCCCGAGGTGCGCGGGCTTGGGAGAGGGCTCTCA 790
Qy 610 -----CAGCCCTGGGGGCACTTCTCAGGCGCAGACAGATTGATTTCCCGTGGCATC 663
Db 791 AGGACGGGGTCTCCAGGATCCGATCCAGCAGAAACCATGATTTGGCTCTGGCCTG 850
Qy 664 CTGGTCCCCCAGTTGTTGGTGGCCATCTCGAAAGAGAGGGCTTAACTTAAGAC 723
Db 851 CTGGTCCCCCAGTTGTTGGAGCCATCAGGAAAGAGGGCCACATTTGGGAAC 910
Qy 724 ATCACTAGCAGACCCAGTCCCGGATAGTATCCATAGAAAGAGAACTTGGAGCTGCA 783
Db 911 ATCAACAAACAGACCCAGTCTAAATGATGTCCATAGAAAGAGAGGGGCTGCT 970
Qy 784 GAGAGGCTGTCACTATCCATCCAGCCAGAGGGAGCTTGAAGATGCGCATGAT 843
Db 971 GAGAGGCTGATTAATCT 1030
Qy 844 CTGAATATCAGCAGAAAGAGGAGATGAGCCAACTAGCCGAAAGATTCCTGAAA 903
Db 1031 CTGAGATTAATGCAATAGAGAGCTCAAGATTAATTAATTCACAGAGAGATCCCTTGAAG 1090
Qy 904 ATCTTGGCACAATAGCTTGGTGAAGATGATGAAAGATGAAAGAGAGAGAAATTGAAG 963
Db 1091 ATTTAGCTCAATTAATCTTGTGGAGCTGTTATTTGTAAGAGAGAGAACTTAA 1150
Qy 964 AAAATTAACATGAACAGGAGCAAGATTAATCAATCTCATCTTTCAGATTTGAGCATA 1023
Db 1151 AAAATTAAGAGAGCAGACACATAAATCAAGATCTCATTTGAGAGATTAAGCCTG 1210
Qy 1024 TACAAACCGGAAAGAACATCATCTGTGAAGGCAAGTTGAGGCTGTGCACTGTGAG 1083
Db 1211 TATATCTCAGAGAGCTATTAACGTTAAAGCATTTGAGACATGTGCAAAAGCTGAG 1270
Qy 1084 ATAGAGATTGAAGAGCTGCGTGAAGCTTTGAAATATGTCGAGCTGTTAAACC 1143
Db 1271 GAGAGATCATGAAAGAAATCAGGAGCTTTTGAATATGATGCTTCTATGAATCTT 1330
Qy 1144 CACTCCGAGT-----ACTTCTCA 1162
Db 1331 CAGACCATTTAATTCCTGATTAATCTGACGCTTGGGTCTGTTCCACCCACTTCA 1390
Qy 1163 GCTGTAACCCCATTCACCAATTTGGCCCTTCCGATCATCTC----- 1208
Db 1391 GGGATGCACTTCCCACTCAGGAGCCCTTCAGCCATGACTCTCCCTACCCGAGTTT 1450
Qy 1209 --TTATCAGAGAGAGATTGTAATCTTTCATCCCAACAGAGTGTGGGCGCAGC 1266
Db 1451 GAGCATATCAGAAAGAGAGATGTTTCACTAGTTTATCCAGCTTATACAGTGGTGCATC 1510
Qy 1267 ATGGGAAGAGGGGCGACATCAACAGCTGGCGAGATTGCGCGAGCTTATCAAG 1326

Db 1511 ATCGCAACAGAGGCGCAGACATCAAGCATTCTCGCTTGGTGGAGCTTCAATTAG 1570
Qy 1327 ATTGCCCTCGGAAAGGCGCAGACATCGACGAAAGATGTCATCATCCGGGCGACCG 1386
Db 1571 ATTGCTCCAGCGGAAAGCAGATGCTAAAGAGATGCTATTACTGACGACCA 1630
Qy 1387 GAGGCCAGTTCAAGGCCAGGAGCGATCTTTGGAAACTGAAGAGAGAACTTCTT 1446
Db 1631 GAGGCTCAGTTCAAGCTCAGGAGAAATTTATGAAAAATTAAGAGAAACTTTGTT 1690
Qy 1447 AACCCAAAGAGAGTGAAGCTGAAAGCGCATTCAGAGTCCCTTCCACAGCTGCG 1506
Db 1691 AGCTTAAGAGAGAGTGAATTTAAGCTCATATCAGATGCTCCATCTTGTGCTGCG 1750
Qy 1507 CGGATGATTGGCAAGGTGCAAGACCGTGAACGATGCAAGCTTAACAGTGCGAGA 1566
Db 1751 AGAGTTATTGGAAGAGAGGCAAAACGTTGAATTAATTTGCAAGTGCAGA 1810
Qy 1567 GTCATGCTGCTCGTACCAAAACGCAATGAAATGAGAAATGATCGTCAAGATTATC 1626
Db 1811 GTTGTGTCCTCGTACCAAGCAGTATGATGAGATGACCAAGTGTGTCAAAATTAATC 1870
Qy 1627 GGGCATTCTTGTAGCAGATGCAAGCGCAAGATCAGGAAATTTGACAAAGTGTG 1686
Db 1871 GGTCACTTATGCTTGGCAGGTGCGCAAGAAATTTCAAGAAATTTCTGACTAGGTA 1930
Qy 1687 AAGCAGCAGAGCAAGAAATACCTCAGGAGTCCGCTCAG 1726
Db 1931 AAGCAGCAGCAACAGAGAGCTCTGCAAGTGGACAC 1970

RESULT 10
US-09-643-597-175
Sequence 175, Application US/09643597
Patent No. 6426072
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Fan, Liqun
APPLICANT: Kalos, Michael D.
APPLICANT: Bangur, Chantanya S.
APPLICANT: Hoeken, Nancy
APPLICANT: Fanger, Gary R.
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aljun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Henderson, Robert A.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.455C11
CURRENT APPLICATION NUMBER: US/09/643,597
CURRENT FILING DATE: 2000-08-21
NUMBER OF SEQ ID NOS: 369
SOFTWARE: PaetsSeq for Windows Version 3.0
SEQ ID NO 175
LENGTH: 4181
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: (3347)
NAME/KEY: unsure
LOCATION: (3502)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (3506)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (3520)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (3538)

OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (3549)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (3646)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (3940)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (3968)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (3974)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (4036)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (4056)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (4062)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (4080)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (4088)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (4115)
 OTHER INFORMATION: n=A,T,C or G
 US-09-643-597-175

Query Match 21.2%; Score 695.6; DB 4; Length 4181;
 Best Local Similarity 64.8%; Pred. No. 2.6e-184;
 Matches 1115; Conservative 0; Mismatches 539; Indels 66; Gaps 3;

73 ATGAAACAGCTTACATCGGGAACCTGAGCCCGCCGTCACCGCGACGACCTCCGCGAG 132
 251 ATGAAACAACTGATATCGGAAACCTCAGGAGAGACGCGCCCTCGGACCTAGAAAGT 310
 133 CTCTTTGGGAGACGAAAGCTGCGGCGGAGCAAGTCTCTGTAAGTCCGCTAGCGC 192
 311 ATCTTCAGAGACGCCAAGATCCCGGTGTCCGGAACCTTCTGTGTAAGTCTGCTAGCG 370
 193 TTGCTGACCTACCGCCGACCGAAGCTGGGCGCATCGCGCCATCGAGACCTCTCGGCTAAA 252
 371 TTGCTGACCTGCGCGGAGCGAGGCTGGGCGCTTCAAGGCGCATCGAGGCGCTTTCAAGTAAA 430
 253 GTGGAATTGCGATGGGAAATCATGGAAGTTGATTACTCACTCTTAAAGCTTAAGAGC 312
 431 ATGAAATCGACGCGGAAACCATTAAGAGTTGAGCACTCGTCCCAAAAAGGCAAGATTT 490
 313 AGGAAATTCAGATTGAAAATCTCTCTCTCACTGAGAGTGGAGGCTTTGAGTGA 372
 491 CGGAAATTCAGATACGAAATATCCGCTCATTTAAGTGGAGGCTGCTGAGTAA 550
 373 TTGCTCAATATGAGGACAGTGGAGATGTGGAACAAGTCAACAGACAGAGAAACCGCC 432
 551 CTATGCTCAGATGAGTGGAGTGGAGAGCTGTGAGCAAGTGAACCTGACTCGGAAACCTGCA 610
 433 GTTGTCAACGTCACATATGCAACAAGAGAGAGCAAAATAGCATGAGAGCTTAAGC 492
 611 GTTGTAAATGTAACCTATTCAGATAGAGACCAAGCTAGACAAAGCACTAAGCAAACTAAT 670
 493 GGGCATAGTTTGAAGACTACTCTTCAAGATTTCTTCAATCCCGGAGTGAAGAGTGAAGC 552
 671 GGAATTCAGTGAAGATTTCACTTGAAGATGACCTATATACCTGTAGTAAATAGCGCGCC 730
 553 TCCCTTGGCCCTCAGCGAGCCAGAGGTGGGAGCACTCTTCCGCGAGCAAGGC-- 609

731 CAGCAAAACCCCTTGACAGAGCCCGAGGCTGCGCGGCGCTTGGAGAGGCGCTCTCA 790
 610 -----CACGCCCCCTGGGGGACCTTTCAGGSCAGACAGATTGATTTCCCGCTGCGATC 663
 791 AGGCAAGGGGTCTCAGAGTCCGTATCCAGAGAGAACATGATGATTTGCTGTGGCTG 850
 664 CTGGTCCCAACCAAGTTGTTGGTCCATCATCGGAAGAGGGCTTAAACATTAAGAAC 723
 851 CTGGTCCCAACCAATTTGTTGGAGCCATCATAGGAAAGAGGGCCACCATTCGGAAC 910
 724 ATCTAAAGCAGACCCAGTCCCGGTAGATATCCAAGAAAAGAACCTGAGAGCTCA 783
 911 ATCAACAAACAGACCCAGCTTAAATGATGATCCACCGTAAAGAAATCGGGGCTGCT 970
 784 GAGAGGCTGTGACCATCATGACCAACCGAGAGGAGCTTTCAGACATGCGCATGAT 843
 971 GAGAGTGCATTAATCTTCTCTTCTGAGGCACTGCGGGCTTGAAGTCTAT 1030
 844 CTGAAATCATGACAGAAAGGAGATGAGACCAATCTAGCCGAAGATTTCTTGAAA 903
 1031 CTGAGATTATGCAATAGGAAAGCTCAGATATTAATCAAGAAAGATCCCTTGAAG 1090
 904 ATCTTGCACACATGCGCTTGTGGAGAGCTGATTTGGAAGAGGAGCAATTTGAAG 963
 1091 ATTTAGCTCATTAATACCTTGTGTGAGCTTTATTTGTAAGAGAGAAATCTTAA 1150
 964 AAATTTGAACATGAACAGGAGCAAGATTAACATCTCATCTTTCAGAGATTGACATA 1023
 1151 AAATTTGAGCAAGACAGACACTAAATCAAGATATCTCATTCAGAGATTGACGCTG 1210
 1024 TACAAACCGGAAAGAACCATCACTGTGAAGGCGACATTTAGGCGCTGTGCGTGA 1083
 1211 TATTAATCCAGAACGCACTATTAAGATTAAAGGCAATGTGAACATGTGCAAGGTGAG 1270
 1084 ATAGAGATTATGAAGAGTGGTGAAGGCTTGAAGATATATGCTGCTGAACAC 1143
 1271 GAGAGATCATGAAGAAATCAAGAGGCTTTATGAAGATGATTTGCTTATGATCTT 1330
 1144 CACTCGGAT-----ACTTCTCA 1162
 1331 CAGGCAATTTAATCTGTGATTAATCTGAAGCCTTGGGTGTCTCCACCACTTCA 1390
 1163 GCGCTGACCCCATCAACGATTTGGCCGCTTCCGATCATCTC----- 1208
 1391 GGGATCCACCTCCACTCAGGCGCCCTTCAAGCATGCTCTTCAACCGCAGTTT 1450
 1209 --TTATCCAGAGAGAGATGTAATCTTTCATCCCAACCGAGCTGTGGCGCATC 1266
 1451 GAGCAATCGAAACGAGAGCTGTCATCACTTTATCCAGCTTATCACTGCTGATC 1510
 1267 ATCGGAAAGAGGAGGACACATCAACAGCTGGAGATTCGCGGAGCTTATCAAG 1326
 1511 ATGCGAGAGAGGCGCAGACATCAAGAGCTTTCGCTTTCGAGCTTCAATTAAG 1570
 1327 ATTGCCCTCGGAGAGGCCCACTGTCAGGAGAAAGATGCTCATCAACCGGCGCACCG 1386
 1571 ATGCTCCAGCGGAGAGCAGATGCTTAAGATGATGATGATATCACTGAGACCA 1630
 1387 GAGGCCAGTTCAAGGCGCAGGAGCGATCTTGGGAACTGAAAGAGAAACTTCTT 1446
 1631 GAGGCTCAGTTCAAGGCTCAGGAGAGATTTATGAAAAATTTAAGAAAGAAACTTGT 1690
 1447 AACCCAAAGAGAGTGAAGCTGGAAGCGCATATCAGAGTCCCTTTCACAGCTGGC 1506
 1691 AGTCTTAAAGAGAGGTAACCTTGAAGCTCATATCAGAGTGCATCTTGTCTGCTGGC 1750
 1507 CGGCTGATGGCAAGGTGCGAAGACCTGTAAGCAATTCAGAGACTTAACAGGCGAGAA 1566
 1751 AGAGTTATGGAAAGAGGAGGCAAAACGTTGATGAATTAAGATTTGTCAAGTGCAGAA 1810
 1567 GTCACTGCTGCTGTCAGAAAGCCAGATGAATGAAGATGATGATGATGATATTC 1626

Db 1811 GTTGTGTCCTCTGACACCACTGATGAGATGACCAAGTGTGTCATAAATGACT 1870
QY 1627 GGGACCTTTTGTCTAGGACACTGACAGCCGACATGAGGAAATTTGACAAAGCTG 1686
Db 1871 GGTCACCTTCTATGCTTGCCAGGTGTCGACGAAATTCAGGAAATTCGACTCAGTA 1930
QY 1687 AAGCAGCAGAGACAGAAATACCTCAGGAGTGCCTCAG 1726
Db 1931 AAGCAGCAGCAGACAGAAAGGCTTGCAAGTGCAGCCAG 1970

RESULT 11

US-09-480-884A-175
Sequence 175, Application US/09480884A
Patent No. 6482597
GENERAL INFORMATION:
APPLICANT: Wang, Tonglong
APPLICANT: Fan, Liqun
APPLICANT: Hosken, Nancy A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fang, Gary R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
FILE REFERENCE: 210121.455C6
CURRENT APPLICATION NUMBER: US/09/480.884A
CURRENT FILING DATE: 2001-08-27
NUMBER OF SEQ ID NOS: 330
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 175
LENGTH: 4181
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(4181)
OTHER INFORMATION: n=A,T,C or G
US-09-480-884A-175

Query Match 21.2%; Score 695.6; DB 4; Length 4181;
Best Local Similarity 64.8%; Pred. No. 2.6e-184;
Matches 115; Conservative 0; Mismatches 539; Indels 66; Gaps 3;

QY 73 ATGACAGAGCTTTATCATCGGAGACCTGAGCCCGCTCACCGGACCTCCGGAG 132
Db 251 ATGACAGAGCTTTATCATCGGAGACCTGAGGAGAACCCCGCTCGACCTTGAAGT 310
QY 133 CTCTTTGGGACAGAGAGCTGCCCCCTGGCGGACAGGTCTGCTGAAGTCGAGTACC 192
Db 311 ATCTTCAAGAGACGCGAAGATCCCGGTGTCGGGACCTTCTGCTGAGAGACTGCTAC 370
QY 193 TTGGTGAAGTACCCGACGAGAACTGGGACATCGGCGCATCGAGACCTCTCGGGTAA 252
Db 371 TTGGTGAAGTACCCGACGAGAGTGGGACCTTCAAGGCACTGAGGGCTTTCAAGTAA 430
QY 253 GTGGAATTCAGTGGGAAATCATGGAATGATTAATCACTCTTAATAAGCTAAGAGC 312
Db 431 ATGGAAGTGAAGGAAACCATGGAATGAGAGCTCGGTCCCAAAAAGCAAGGAT 490
QY 313 AGGAAATTCAGTGGAAATCATGCTCTCTCTCACTGAGTGGAGAGTGTGATGAGCTT 372
Db 491 CGGAACTTCAAGTGAAGAAATATCCGCTCATTTTCAAGTGGAGGCTGAGTAACTTAA 550
QY 373 TTGGCTCAATATGAGGACAGTGAAGTGTGGAACAGTCAACACAGACAGAAACCGCC 432
Db 551 CTAGTCCAGTATGAGAGTGGTGAAGAGTGTGAGCAAGTGAACCTGACTGGAAACTGCA 610
QY 433 GTTGTCAAGCTCAATATGCAACAGAGAGAGAAATAAGCTAGAGAGTAAAG 492
Db 611 GTTGTAAATGTAACCTATTCAGTAAAGCAAGGCTAGCAAGCACTAAGCAACTGAAT 670
QY 493 GGGCATTCAGTGTAGAACTACTCTTCAAGATTTCTTCACTCCCGAGTGAAGAGTGAAG 552
Db 671 GGATTTCAAGTGAAGATTTCACTTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 730

QY 553 TCCCTTGCCTCCCTTACGAGACCCAGCGTGGGAGCACTCTTCCGGAGAGAGC--- 609
Db 731 CAGCAAAACCCCTTCCAGAGCCCGAGGTCGCCCGGGGCTTGCGGAGAGGGGCTCTCA 790
QY 610 -----CAGGCCCTGGGGGACCTTCTCAGGCAAGACATTTGATTTCCCGTGGGATC 663
Db 791 AAGCAGGGGCTTCAGAGATCCGTATCCAGGAGAAACATGATTTGCTCTGCGCCTG 850
QY 664 CTGATCCCAAGCAGTTTGTGTGCTCATCGGAAAGAGGGCTGACCTTAAGAAC 723
Db 851 CTGATCCCAAGCAGTTTGTGTGCTCATCGGAAAGAGGGCTGACCTTAAGAAC 910
QY 724 ATCTAAGCAGAGCCAGTCCCGGTAGATATCTAAGAAAGAAAGAACTCTGAGCTGCA 783
Db 911 ATCAACAAACAGACCCAGCTTAAATTCAGTGTCCACCGTAAGAAATAGCGGGGCTGCT 970
QY 784 GAGAGCGGTGACCATTCATGCGACCCCAAGAGGAGCTTGAAGCATGCGCATGAT 843
Db 971 GAGAGGTGATTAATCTCTCTTACTCTGTAAGGACCTCTGCGCTTGAAGTCTAAT 1030
QY 844 CTGAAATCATGAGAAAGAGGAGATGAGACCAACTAGCCGAAAGATTCCTGAAA 903
Db 1031 CTGAGATTAATGCAATAGGAGCTCAGATATTAATTCAGAGAGAGATCCCTTGAAG 1090
QY 904 ATCTTGGCAGCAATGCTTGGTGGAGAGTGAATTTGAGAAAGAGGAGAAATTTGAAG 963
Db 1091 ATTTAGCTCAATTAATCTTTGTTGAGAGTCTTATTTGTTAAGAAAGAAATTTTAA 1150
QY 964 AAAATTAAGCAATGAAACAGGAGCAAGTAATCATCTTTGACAGATTTAGCAT 1023
Db 1151 AAAATTAAGCAAGCAAGCAAGCAATTAATCATCTTTGACAGATTTAGCAT 1210
QY 1024 TACACCCGAGAAAGCAATCATCTGTGAGGAGCAGTGTGAGGCTGCGCAGCTGAG 1083
Db 1211 TATATCAGAGAGCACTATTCAGTTTAAAGCAATTTGAGAGATGTCGCAAGCTGAG 1270
QY 1084 ATGAGATTAATGAAAGAGCTGCGAGGCTTTGAAATGATGCTGCTTAAACAC 1143
Db 1271 GAGGAGATTAATGAAAGAGCTGAGGAGCTTATGAAATGATTTGCTTATGATCTT 1330
QY 1144 CACTCCGGAT-----ACTTTCGA 1162
Db 1331 CAGGACATTAATTTCTGATTAATTCGAGCTTGGGCTGTTCCACCACTTCA 1390
QY 1163 GCTGTACCCCATCAAGATTTGGCCGTTCCCGCATCACTC----- 1208
Db 1391 GGGATGCACTTCCCACTCAGGAGCCCTTTCAGCCATGACTCTCTTACCCGCACTT 1450
QY 1209 --TTATCAGAGAGAGATTTGATCTCTTATCCCAAGCCAGGCTGTGGGCGCATC 1266
Db 1451 GAGCAATCAGAAACGAGAGCTGTTCACTGATTAATCCAGCTCATAGTGGGCTATC 1510
QY 1267 ATGCGAAGAGAGGGGAGCATCAATCAAGCTGCGGAGATTCGCGGAGCTTATCAAG 1326
Db 1511 ATGCGAAGAGAGGAGAGCATCAAGCACTTCTCGCTTGTGAGGCTTCAATTAAG 1570
QY 1327 ATTGCCCTTGGGAGGCGCAGAGCTCAGGAGAAAGTGTGATCATCAAGGCGCAGCG 1386
Db 1571 ATTGCCCTTGGGAGGAG 1630
QY 1387 GAAAGCCAGTTCAGGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1446
Db 1631 GAGGCTCAGTTCAAGGCTCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1690
QY 1447 AACCCCAAG 1506
Db 1691 AGTCTTAAG 1750
QY 1507 CGGATGATTTGAG 1566
Db 1751 AGAGTTATTTGAG 1810

QY	1567	GTCATCGTGCCTCGTGAACCAAGCCGAGATGAATAATGAGGAAGATGCTGAGAAATATC	1626
Db	1811	GTGTGTTCCTCCGTGACACGAGACCTCGATGAGATGACCAAGTGTTGTAAATAATCT	1870
QY	1627	GGGCACTTCTTTGGTAGCCGAGACGACAGGCAAGATCAGGAAATGTGACCAAGGTG	1686
Db	1871	GGTACTCTTATGTGTTCAGAGTGGCCGAGGAAAAATTCAGGAAATTTCTAGTCAGSTA	1930
QY	1687	AAGCAGCAGAGCAGAAATACCTTCAGGAGATCGCCTCAC	1726
Db	1931	AAGCAGCACCACCAACAGAGAGGCTCTGCAAAAGTGAACAC	1970

```

1 RESULT 12
2 US-09-542-615A-175
3 ; Sequence 175, Application US/09542615A
4 Patent No. 6518256
5 GENERAL INFORMATION:
6 APPLICANT: Wang, Tongtong
7 APPLICANT: Fan, Liqun
8 APPLICANT: Kalos, Michael D.
9 APPLICANT: Bangur, Chaltanya S.
10 APPLICANT: Hosken, Nancy A.
11 APPLICANT: Fanger, Gary R.
12 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
13 TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
14 FILE REFERENCE: 210121,45568
15 CURRENT APPLICATION NUMBER: US/09/542,615A
16 CURRENT FILING DATE: 2000-04-14
17 NUMBER OF SEQ ID NOS: 350
18 SOFTWARE: FastSeq for Windows Version 3.0
19 SEQ ID NO 175
20 LENGTH: 4181
21 TYPE: DNA
22 ORGANISM: Homo sapiens
23 FEATURES:
24 NAME/KEY: unsure
25 LOCATION: (3347)
26 OTHER INFORMATION: n=A,T,C or G
27 NAME/KEY: unsure
28 LOCATION: (3502)
29 OTHER INFORMATION: n=A,T,C or G
30 NAME/KEY: unsure
31 LOCATION: (3506)
32 OTHER INFORMATION: n=A,T,C or G
33 NAME/KEY: unsure
34 LOCATION: (3520)
35 OTHER INFORMATION: n=A,T,C or G
36 NAME/KEY: unsure
37 LOCATION: (3538)
38 OTHER INFORMATION: n=A,T,C or G
39 NAME/KEY: unsure
40 LOCATION: (3549)
41 OTHER INFORMATION: n=A,T,C or G
42 NAME/KEY: unsure
43 LOCATION: (3646)
44 OTHER INFORMATION: n=A,T,C or G
45 NAME/KEY: unsure
46 LOCATION: (3940)
47 OTHER INFORMATION: n=A,T,C or G
48 NAME/KEY: unsure
49 LOCATION: (3968)
50 OTHER INFORMATION: n=A,T,C or G
51 NAME/KEY: unsure
52 LOCATION: (3974)
53 OTHER INFORMATION: n=A,T,C or G
54 NAME/KEY: unsure
55 LOCATION: (4036)
56 OTHER INFORMATION: n=A,T,C or G
57 NAME/KEY: unsure
58 LOCATION: (4056)
59 OTHER INFORMATION: n=A,T,C or G
60 NAME/KEY: unsure

```

```

1 LOCATION: (4062)
2 OTHER INFORMATION: n=A, T, C or G
3 NAME/KEY: unsure
4 LOCATION: (4080)
5 OTHER INFORMATION: n=A, T, C or G
6 NAME/KEY: unsure
7 LOCATION: (4088)
8 OTHER INFORMATION: n=A, T, C or G
9 NAME/KEY: unsure
10 LOCATION: (4115)
11 OTHER INFORMATION: n=A, T, C or G
12
13 US-09-542-615A-175

```

Query Match	21.2%	Score 695.61	DB 4	Length 4181
Best Local Similarity	64.8%	Pred. No. 2.6e-184		
Matches 1115	Conservative	0	Mismatches 539	Indels 66
				Gaps 3
QY	73	ATGAAACAAGCTTTATCAATCGGGAACCTGAGCCCGCGCTGACCGCGAGACGACTCCGGCAG	132	
Db	251	ATGAAACAACCTGATATATTCGGAAACCTCAGCGAGAAAGCGCGCCCTCGAGCTAGAAAGT	310	
QY	133	CTCTTTGGGGACAGAAAGCTGCGCCCTGGGGGAGACAGTCTGCTGAAGTCGGCTACGCC	192	
Db	311	ATCTTCAAAGACCCCAAGATCCGGGTGTGGGACCTTCTGTGTAAAGCTGGCTACGGG	370	
QY	193	TTGCTGACACTACCCGACACAGAACTGCGGCATCCGCGCATCGAGACCTCTCGGGTAA	252	
Db	371	TTCTGTGATGCTCCCGACGAGAGCTGGGCGCTCAAGGCATCGAGGCGCTTTCAAGTAA	430	
QY	253	GTGGAATTGCAATGGGAATAATCATGGAAGTTGATTACTAGTCTCTAAAGTAAGAGAC	312	
Db	431	ATAGAACTCGACGGGAAACCCATAGAAATTGAGCACTCGTCCAAAAAGCAAGAGATT	490	
QY	313	AGGAAATTTGAGATTGCAACATCCCTCCTCACTGAGTGGAGGTGTTGATGACTT	372	
Db	491	CGAAATCTTGATACGAATATTCGCGCTCATTTTCAGTGGGAGGTGCTGATAGTTTA	550	
QY	373	TTGGCTCAATATGGGACAGTGGAGATGTGGAACTGGAACAGTCAACACACAGAACCGGC	432	
Db	551	CTAGTCAGATGTGAGAGTGTGGAGAGCTGTGAGCAATGTGAACACTGACTCGGAACTGCA	610	
QY	433	GTTGTCAACGTCAATATGCAACAAGAAAGAAACAAAATACCCATGAGAAAGCTTAAG	492	
Db	611	GTTGAAATGTAACTTATTCAGTAAGAACCAAGCTTGACAAACACTAGCAAACTGAAT	670	
QY	493	GGGCACTCAGTTTGAAGAACTACTCTCTTCAAGATTTCTTAATCCCGGATGAAGGTGAGC	552	
Db	671	GGATTTGAGTTAGAAATTTCACTTGAAAGTAACTTATTCCTGATGAATAGGCGCGC	730	
QY	553	TCCCTTTGCCCTCTCAGCGAGCCACAGGTGGGAGACACTTCCCGGAGCAAGGC---	609	
Db	731	CAGCAAAACCCCTTGACAGACCCCGAGAGTGC CGGGGGCTTGGGACAGAGGGCTCTCA	790	
QY	610	-----CAGCGCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCGCTGCGATC	663	
Db	791	AGCAGGGGCTCTCCAGATTCGATATCCAGCAGAAACCATGTGATTTGGCTCTGCGGCTG	850	
QY	664	CTGTGTCCCAACCCAGTGTGTGTGGCCATCATTCGAAAGAGGGCTTACCATTAAGAC	723	
Db	851	CTGTGTCCCAACCCAAATTTGTGAGCCATCATAGGAAAGAGGTGCACCATTTGGAAC	910	
QY	724	ATCATCAAGCAGACCCAGTCCGGGCTAGATATTCATAGAAAGAAACTTGTGAGCTGCA	783	
Db	911	ATCACCAACAGACCCAGTCTAAATTCATATGATGCCACGTAAAGAAATGCGGGGGCTGCT	970	
QY	784	GAGAACCTGTCAACATCCATGCCACCCACAGGGGAACTTCTGAAGCATGCGCATGTT	843	
Db	971	GAGAACTGCATTAATCAATCTCTCTACTCTGTAAAGCACTTGCGGCTTTGAATCTAT	1030	
QY	844	CTTGAATATATCAGAAAGGCAAGCAGATGAGACCAACTAGCCGAGAGATTTCTTGAA	903	
Db	1031	CTGGAATATATGATATGAGAACTCAAGATATATAATTTCAAGAGAGATGCCCTTGAAG	1090	

QY 904 ATCTGGCACAATGGCTTGGTTGGAGACTGATTGGAAAAAGAGCAGAAATTTGAAG 963
Db 1091 ATTTAGCTCATTAATACTTTTGTGACGCTTTATTTGTAAGAAAGAAATCTTAAA 1150
QY 964 AAAATTTGAACATGAACAGGGGCCAAGATTAACAATCTACTTTTGAGAGATTGACATA 1023
Db 1151 AAAATTTGACAAAGACACAGACTAAATACAGATATCTCCATTCAGAGAAATTCAGCTG 1210
QY 1024 TACAACCCGGAAGAACCATCATCTGTGAAGGGCAGAGTGGAGCCGTGGCCAGTCTGAG 1083
Db 1211 TATAATCCAGAAACGACATTTACAGTTAAAGGCATGTTTGAACATGTGCCAAGAGCTGAG 1270
QY 1084 ATAGAGATTATGAAGAGCTGCTGAGGCTTTGAAAAATGATATGCTGTTAAACACC 1143
Db 1271 GAGAGATCATGAAGAAATCAGGAGTCTTATGAAATGATATGCTTATGAAATCTT 1330
QY 1144 CACTCCGGAT-----ACTTGTCCA 1162
Db 1331 CAAGCACTTTAATCTCGATTAATCTGAACGCTTGGTCTGTTCCACCCACTTCA 1390
QY 1163 GCTGTATACCCCATCACCAAGTTTGGCCCGCTCCCGCATCATCATC----- 1208
Db 1391 GGGATGCCACTCCCACTCTCAGGGCCCCCTTCAGCATGACTCTCCCTACCCGCAAGTTT 1450
QY 1209 --TTATCAGACGACGAGATTTGTAATCTTCAATCCCAACCCAGGCTGTGGCGCATC 1286
Db 1451 GAGCAATCAGAAACGAGAGACTGTCATCAGTTTATCCAGCTCATAGTCGGTCCATC 1510
QY 1267 ATCGGGAAGAAAGGGGGCACAACATTAACAGCTGCGAGATTTGCGCGAGCTTATCAAG 1326
Db 1511 ATCGGCAAGCAGGGGCGACACATCAGCAGCTTTCTGCTTGGTGAAGCTTCAATTAAG 1570
QY 1327 ATTGCCCCGCGGAAGGCGGAGCGTCAAGCGAAGAGATGTCATCATCACCGGGCCACCG 1386
Db 1571 ATTGCTCAGCGGAGACCAAGATGCTTAAAGTGAAGATGATTTACATGACCAACA 1630
QY 1387 GAAGCCCAATTCAAGGCCACGAGCGATCTTTGGGAAACGAAAGAGAAATCTTCTT 1446
Db 1631 GAGGCTCAGTTCAAGGCTCAGGGAAGATTTATGAAAAATTAAGAAAGAAATCTTGT 1690
QY 1447 AACCCCAAGAAAGAGGAGCTGGAAGCGCATATCAAGTGCCTTTCACAGCTGGC 1506
Db 1691 AGTCTTAAAGAAAGGGTGAACCTTGAAGCTCATATCAGATGCTTGTGCTGGCG 1750
QY 1507 CGGGTGAATGGCAAGGTGGCAAGACCGTGAACGAACTGAGAACTTAACAGTGCAGAA 1566
Db 1751 AGAGTTATTGAAAAAGAGCAAAACGTTGATGAATCTTCAAGATTTGTCAAGTGCAGAA 1810
QY 1567 GTCATCGTGCCTGTGACCAACGCGCATGAATGAATGAGAAATGATCTCAGAAATATC 1626
Db 1811 GTTGTGTGCTCGTGTGACGACACCTGATGAGATGACCAAGTGTGTCAAAATTAAT 1870
QY 1627 GGGCACTTCTTCTGAGCCGACATGCAACGCGCAAGATCAGGAAATTTGTAACAAGGTG 1686
Db 1871 GGTCACTTCTATGCTTGGCCAGGTTGCCAGAGAAAAATTCAGAAATTCGACTCAGGTA 1930
QY 1687 AAGCAGCAGAGAGAGAAATACCTTCAAGGAGTGCCTCAGC 1726
Db 1931 AAGCAGCAGCAGAGAGAGGCTTGCAGAAATGAGCACAC 1970

RESULT 13
US-09-606-421B-175

; Sequence 175, Application US/09606421B
; Patent No. 6531315

; GENERAL INFORMATION:

; APPLICANT: Wang, Tonglong

; APPLICANT: Fan, Liqun

; APPLICANT: Kalos, Michael D.

; APPLICANT: Bangur, Chaitanya S.

; APPLICANT: Hosken, Nancy

; APPLICANT: Fanger, Gary R.

; APPLICANT: Li, Samuel X.

; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Jasir A.W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.455C9
; CURRENT APPLICATION NUMBER: US/09/606,421B
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 175
; LENGTH: 4181
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: unsure
; LOCATION: (3347)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3502)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3506)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3520)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3549)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3646)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3940)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3968)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3974)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (4036)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (4056)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (4080)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (4088)
; OTHER INFORMATION: n=A,T,C or G

FEATURE:
 NAME/KEY: unsure
 LOCATION: (4115)
 OTHER INFORMATION: n=A,T,C or G
 US-09-606-421B-175

Query Match 21.2%; Score 695.6; DB 4; Length 4181;
 Best Local Similarity 64.8%; Pred. No. 2.6e-184;
 Matches 1115; Conservative 0; Mismatches 539; Indels 66; Gaps 3;

```

QY 73 ATGAACAAGCTTTCATCGGGAACCTGAGCCCGCCCGTCAACCGCCGAGACGACCTCCGCGAG 132
DB 251 ATGAACAAGCTTTCATCGGGAACCTGAGCCCGCCCGTCAACCGCCGAGACGACCTCCGCGAG 310
QY 133 CTCTTTGGGGAACAGAAAGCTGCCCCCTGGGGAACAGTCTGCTGAAGTCCGGCTAGCC 192
DB 311 ATCTTCAAGAGCGCCAGATCCCGGTCGAGACCTCTCTGGAAGTGGCTACCGC 370
QY 193 TTCTGGAACCTCCCGACCAAGCTGGGGCAATCCGCGCATCGAGACCTCTCGGGTAA 252
DB 371 TTCTGGAACCTCCCGACCAAGCTGGGGCAATCCGCGCATCGAGACCTCTCGGGTAA 430
QY 253 GTGAATTCAGTGGGAATATCATGGAAGTTGATTAATCATGCTCTTAAAGCTAAGAGC 312
DB 431 ATGAGACTGACGAGAAACCATAGAGTTGAGCACTCGGTCCCAAAAGGCAAGGATT 490
QY 313 AGAAATTCAGATTGAAACATCCCTCTCTCACTGCACTGGAGAGTGTGGATGAGCTT 372
DB 491 CGAAATCTTCAAGATCAGAAATATCCCGCTCATTTACAGTGGAGAGTGTGGATGAGCTT 550
QY 373 TTGCTCATATGGAACAGTGGAGAAATGGAACAAGTCAACAGACACAGAAACCGCC 432
DB 551 CTAGTCCAGTATGAGAGTGGTGGAGCTGTGCAAGTGAACCTGACTGACCTCGAACTGCA 610
QY 433 GTTGTCAACGTACATATGCAACAAGAGAGAGAAATATGCAATGAGAGAGTAAAGC 492
DB 611 GTTGTAAATGTAACTATTTCCAGTAAAGACCAAGCTAGCAAGCACTAAGCAAACTGAT 670
QY 493 GGGCATCAGTTTGAAGTACTCTCTTCAAGATTCTTCAATCCCGATGAAGAGTGAAGC 552
DB 671 GGAATTCAGTAAAGATTTCACTTGAAGAGTAAAGTAACTATATCCGTATAAATGAGCGCC 730
QY 553 TCCCTCTGCCCCCTCAGGAGGCCAGCGTGGGGAACACTTTCCGGGAGCAAGCG-- 609
DB 731 CAGCAAAACCCCTTGAAGAGCCCGAGGTCGCGCGGGGCTGGGAGAGGGGCTCTCA 790
QY 610 -----CAGCCCTTGGGGGCACTTCTCAGGCCGACAGATTGATTTCCGCTGGGATC 663
DB 791 AGGCAAGGGTCTTCCAGGATCCGTATCCAGCAAGAAACCATGATTTGCTGCGCTG 850
QY 664 CTGGTCCCAACCCGATTTGTTGTCATCATCGAAAGAGAGGGCTTGACCAATAAGAAC 723
DB 851 CTGGTCCCAACCCGATTTGTTGTCATCATCGAAAGAGAGGGCTTGACCAATAAGAAC 910
QY 724 ATGACTAAGAGAGCCAGTCCCGGATATCATATAGAAAGAGAACTCTGAGCTGCA 783
DB 911 ATGACCAAGAGAGCCAGTATTAATGATGATCCACCGTAAAGAAATGCGGGGCTGCT 970
QY 784 GAGAGCTGTACATCATCATGACCCCAAGAGGGAATTTCTGAAGCATGCCCATGATT 843
DB 971 GAGAGCTGTACATCATCATGACCCCAAGAGGGAATTTCTGAAGCATGCCCATGATT 1030
QY 844 CTGGAATTCAGAGAAAGGAGATGAGACCAAACTAGCCGAAAGAGATTTCTCTGAAA 903
DB 1031 CTGGAATTCAGAGAAAGGAGATGAGACCAAACTAGCCGAAAGAGATTTCTCTGAAA 1090
QY 904 ATCTGGCAACACGCTGTTGTTGAGAGCTGTTGAGAAAGAGGCAAAATTTGAG 963
DB 1091 ATTTAGCTATATATACTTTGTTGAGAGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1150
QY 964 AAAATTGACATGAAAAGAGGAGCAAGATTAATCTCATCTTTGAGAGATTGAGATA 1023
DB 1151 AAAATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1210
  
```

```

QY 1024 TACAAACCCGGAAGAAACCATCATCTGTGAAGGCAACAGTTGAGGCTTGCCAGTCTGAG 1083
DB 1211 TATATTCAGAGAGCACTATTAACGTTAAAGGACATGTTGAGCATGTGCGAAAGCTGAG 1270
QY 1084 ATGAGATTAATGAAGAGCTGCTGAGGCTTTTAAATATGATGCTGCGCTTAAACACC 1143
DB 1271 GAGGAGATCATGAAGAAATCAGGAGAGTCTTATTAATGATGATTTGCTTATATATCTT 1330
QY 1144 CACTCCGGAT-----ACTTCTCA 1162
DB 1331 CAGCAATTTAATTCCTGATTAATCTGAAGCCTTGGGTCTGTTCCACCCACTTCA 1390
QY 1163 GCCTGTACCCCATCACAGTTTGGCCGTTCCCGCATCATCTC----- 1208
DB 1391 GGAATGCACTCCCATCCTCAGGAGCCCTTCAACCATGACTCTCCCTTACCCGAGTTT 1450
QY 1209 --TTATCAGAGCAGAGATTTGATCTCTTCAATCCCAACCAAGGCTGTGGGCGCATC 1266
DB 1451 GAGCAATCAGAAAGAGAGCTGTTATCATGTTATCCAGCTCTATCATGCTGATGCTGATC 1510
QY 1267 ATCGGAAGAGAGGAGGAGCAGATCAACAGCTGGCAGATTTGCGGAGCCTTATCAAG 1326
DB 1511 ATCGGAGAGAGGAGGAGCAGATCAACAGCTTTCTGCTTGTGAGCTTCAATTAAG 1570
QY 1327 ATTGCCCCGGAAGGCGCAGAGCTCAGCGAAGAGATGTCATCATCAACCGGCGCACCG 1386
DB 1571 ATTGCTCAGGAGAGCAGCATGATCTAAAGTGAAGATGATGATTAATCATGAGACACCA 1630
QY 1387 GAGGCCGCTTCAAGGCCAGGAGAGGATCTTTGGGAAACGTAAGAGAGAAATCTCTTT 1446
DB 1631 GAGGCTCAGTTCAAGGCTCAGGAGAAATTTATGAAATTAATTAAGAGAGAAATCTTTGT 1690
QY 1447 AACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1506
DB 1691 AGCTCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1750
QY 1507 CGGATGATTTGCAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1566
DB 1751 AGAGTTATTTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1810
QY 1567 GTCATGTCGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1626
DB 1811 GTTGTGTCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1870
QY 1627 GGGCACTTCTTGTAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1686
DB 1871 GGTCACTTCTATGCTTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1930
QY 1687 AAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1726
DB 1931 AAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1970
  
```

RESULT 14
 US-09-261-855-1
 ; Sequence 1, Application US/09261855A
 ; Patent No. 6255055
 ; GENERAL INFORMATION:
 ; APPLICANT: Ross, Jeffrey
 ; TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN
 ; FILE REFERENCE: 960296.95131
 ; CURRENT APPLICATION NUMBER: US/09/261, 855A
 ; CURRENT FILING DATE: 1999-03-03
 ; NUMBER OF SEQ ID NOS: 46
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 2224
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 US-09-261-855-1

Query Match 20.9%; Score 685; DB 3; Length 2224;
 Best Local Similarity 64.1%; Pred. No. 1.5e-181;
 Matches 1117; Conservative 0; Mismatches 560; Indels 66; Gaps 3;

QY 73 ATGACAAGCTTTAATCGGGAACCTGAGCCCGCCGCTACCGCCGACGACTCCGCGAC 132
 DB 131 ATGACAAGCTTTAATCGGGAACCTGAGAGTGTGACCCCGCAGACTTTGAGAAA 190
 QY 133 CTCTTTGGGGACAGGAGCTGCCCCGCGGGACAGGCTCTGTGAGTCCGGCTACGCC 192
 DB 191 GTATTCCGGGACACAGATCTCCCTACAGCGCCGCTTTGTGTAAATCCGGCTACGCC 250
 QY 193 TTCTGGACCTACCCCGACCAAGACTGGGCCATCCGGCCATCGAACCCCTCTCGGTTAA 252
 DB 251 TTCTGGACCTACCCCGACCAAGACTGGGCCATCGAACCCCTCTCGGTTAA 310
 QY 253 GTGGAATTCGATGGGAAAATATGAGAACTGTTACTAGTCTCTAAAAGCTAAGAC 312
 DB 311 GTAGAACTGCAAGGAAAAGCTGTAGAGATTGAACTCAGTCCCAAAAAAATAAGGAGT 370
 QY 313 AGGAAATTCAGATTTCGAATTCGATCCCTCTCAGCTGCGAGTGGGAGTGTGTGATGACTT 372
 DB 371 CGGAAATTCAGATTCGGAATTCCTCAGCTCCGATGGGAAAGCTAGATGAGCTG 430
 QY 373 TTGGCTCAATTTGGGACAGTGGAGATGTGGAACAGTCAACAGACACAGAAACGCGC 432
 DB 431 CTGGCTCAATTTGGGACAGTGGAGATGTGGAACAGTCAACAGACACAGAAACGCGC 490
 QY 433 GTTGTCAACGCTCATATGCAACAAGAGAAAGCAAAAAATAGCTATGAGAGCTAAGC 492
 DB 491 GTTGTCAACGCTCATATGCAACAAGAGAAAGCAAAAAATAGCTATGAGAGCTAAGT 550
 QY 493 GGGCATCAGTTTGAAGTCTACTCTTCAAGATTCTCAATCCCGGATGGAAGAGTGAAC 552
 DB 551 GGGCATCAGTTTGAAGTCTACTCTTCAATCCCGGATGGAAGAGTGAAC 610
 QY 553 TCCCCCTC-----GCCCCCTCAGCGAGCCGAGCGTGGGAGCACTCTCCCGGAG 603
 DB 611 CAAGTCTCTGAGATGGGCGTCTGAGAGCTTTGGGCTCTGGGGCCAGCCCGGCAAGG 670
 QY 604 CAAGGCGAGGCGCTGGGGGCACTTCTCAGGCCAGACAGATTATTCCTCGCTGGGATC 663
 DB 671 TCGCCCGTGGCAGAGGGGCTCCAGCCAGAGAGCAGTGGGACATCCCTCTCCGGCTC 720
 QY 664 CTGGTCCCGCAGCTTTGTTGGTCCATCATCGGAAAGGAGGGCTTGAACATTAAGAAC 723
 DB 731 CTGGTCCCGCAGAGTGTGAGGCGCTATATTTGGCAAGGAGGCTGCACCTATCCGAAC 790
 QY 724 ATCACTAAGCAGACCCAGTCCCGGTAATATCATAGAAAAGAACTTGGAGCTGCA 783
 DB 791 ATCACTAAGCAGACCCAGTCCCGGTAATATCATAGAAAAGAAATGGGCGCTGCG 850
 QY 784 GAGAGCTCTGTCACTATCCATGCTCAAGAGGAGGACTTCTGAAGCATCGCGATATT 843
 DB 851 GAGAGGCGCTCATCGCTGATTTAAACCCCTGAAGGCTGCTCTCCGCTGCAAGATGATC 910
 QY 844 CTGGAATCATGAGAAAAGGAGCAGATGAGACCAAACTAGCCGAGAGATTCCTCTGAAA 903
 DB 911 TTGGAATTTATGACAAAGAGCAAAAGACACAAAACGCAATGAAGTTCCCTGGAAG 970
 QY 904 ATCTTGGCACAACATGCTTTGTTGGAAGACTGATTGAAAAGGAGCAAAATTTGAAG 963
 DB 971 ATCTTGGCTCATTAACAATTCGTGGGCGACTCATTTGGCAAGAAAGCCGGAACCTGAA 1030
 QY 964 AAAATTTGAACATGAACAGGAGCAAGATTAACAATTCATCTTTGAGAGATTTGACATA 1023
 DB 1031 AAGGTGGACAGAGCAACAGAGCAAGATTAACAATTCATCTGCTCAGAGCTCAGGCTC 1090
 QY 1024 TACAACCCCGGAAAGAACCTCACTGTGAAGGGGACAGTTGAGGCTGTGCTGCTGTGAG 1083
 DB 1091 TATTAACCTCGAGAGGACCATCACTGTGAAGGGGCGCATTTGAGAACTGTTGACAGGCGGAG 1150
 QY 1084 ATTAGAATTATGAAGAGCTGCGTGAAGGCTTTGAAAATGATATGCTGCTGTTAACACC 1143

DB 1151 CAGAGATCATGAAGAAAGTTGTGAGAGGCTTTACGAGAACGAGCTGGCCGCAATGAGCTTG 1210
 QY 1144 CACTCCGATACCTTCTCCAGCCTGTACC-----CCC 1174
 DB 1211 CAGTCCAGCTCATCTCCGGGCTTAACCTGGCTGTGTAGTCTTCTCCAGCTTCATCC 1270
 QY 1175 ATCAACAGTTTGGCCCTTCCCGATCACT----- 1207
 DB 1271 AGCGTGTCCCTCTCTCCAGAGTGTCACTGGGAGCTGTCCCTATATGCTCTTATG 1330
 QY 1208 -CTTATCCAGAGAGAGATTTGATATCTCTTATCCCAACCCAGGCTGTGGGCGCATC 1266
 DB 1331 CAGGCTCCGGAGCAGAGATGTACAACTGTTCATCCCGCCAGGCTGTGGGCCCATC 1390
 QY 1267 ATCGGAAAGAGAGGGGCGACACATATAACAGCTGCGAGATTTGCGCGGAGCCTTATCAAG 1326
 DB 1391 ATTGGCAAGAGAGGCGCAGACATATAAACATCTCCGCTTTCGCGAGGCTTCATCAAG 1450
 QY 1327 ATTGCCCCCTCGGAAAGCCCGAGAGCTACGCAAGAAAGTGTATCATCATCACCGGCAACG 1386
 DB 1451 ATTGCTCCAGCAAGAACCACTGACTCAAGATTGGAATGTGTGATCATCTGACCCCA 1510
 QY 1387 GAAGCCAGTTCAAGGCCCGAGGACGATCTTTGGGAACTGAAAGAGAAAATTCTTT 1446
 DB 1511 GAGGCTAAGTTCAAGGCTCGAGGAAAGATTTATGGCAAACTTAAAGAAAGAAATTTCTT 1570
 QY 1447 AACCCAAAGAGAAAGTGAAGCTGGAAGCGCATATCAGAGTGGCCCTTCCAGAGCGGC 1506
 DB 1571 GGTCCAAAGAGAGAAAGTAAAGCTAGAGACCCACATACGAGGTTCCGCTTTCAGAGCGGC 1630
 QY 1507 CGGATGTTGGCAAGGCTGCAAGACCGTGAACCACTGAGAACTTAACCACTGACAGAA 1566
 DB 1631 CGGTCATTCGCGCAAGAGCGGCAAAACGTTGAATGAGCTGCAGAACTTGACGAGCTGAG 1690
 QY 1567 GTCATGTGCTCTGTGACCAAAAGCCGATGAAATATGAGAAAGTGAATCGTGAATATTC 1626
 DB 1691 GTGATGTGCGCAAGAGCCAGACCCCGATGAGAAACCAAGCATTTGTTAATATATC 1750
 QY 1627 GGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAATTTGTAACAACAGGTG 1686
 DB 1751 GGAATTTTATTCAGCCAGATGAGCTCAGCGGAAGATCCGAACATCTCGGCTCAAGTT 1810
 QY 1687 AAGCAGCAGAGCAAGAAATACCTCAGAGAGTGCCTCAGAGCGCAGCAAGTGAAGCTCC 1746
 DB 1811 AAGCAACAGCACCAAGAGGAGCAGAGCAACCTGCGCCAGGCAAGAGATGACCCGCG 1870
 QY 1747 CAC 1749
 DB 1871 CCC 1873

RESULT 15
 US-09-061-709-7
 ; Sequence 7, Application US/09061709B
 ; Patent No. 6297364
 ; GENERAL INFORMATION:
 ; APPLICANT: Chen, Yao-Tseng
 ; APPLICANT: Gure, Ali
 ; APPLICANT: Tseng, Solam
 ; APPLICANT: Stockert, Elisabeth
 ; APPLICANT: Jager, Elke
 ; APPLICANT: Knuth, Alexander
 ; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
 ; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
 ; FILE REFERENCE: LUD 5538
 ; CURRENT APPLICATION NUMBER: US/09/061,709B
 ; CURRENT FILING DATE: 1998-04-17
 ; SEQ ID NO 7
 ; LENGTH: 1946
 ; TYPE: DNA

ORGANISM: Homo sapiens
FEATURE:

Query Match 17.8%; Score 585.4; DB 3; Length 1946;
Best Local Similarity 64.1%; Pred. No. 1.2e-153;
Matches 964; Conservative 0; Mismatches 481; Indels 60; Gaps 3;

303 GCTAAGAGCAGAAATTCAGATTGCAACATCCCTCTCTCACTGCGAGTGGAGTCTT 362
104 GGTTAGAGAGCCGGAATTCAGAAATTCAGAAATTCAGAAATTCAGAAATTCAGAAATTC 163
363 GGTAG 422
164 GAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 223
423 AGAAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 482
224 TGAAGCAG 283
483 GAAAGTAAAG 542
284 GAAAGTAAAG 343
543 -----AAGAGTAAAG 593
344 GCAGATAGCAG 403
594 TTCCGAG 653
404 CCGCAG 463
654 GCTGAG 713
464 CTTGAG 523
714 CATTAAG 773
524 CATCCGAG 583
774 TGGAGTGCAG 833
584 AGGTGAG 643
834 CCGCATGATCTTGAATATGATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 893
644 TAAGATGATCTTGAATATGATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 703
894 TCTCTGAAG 953
704 TCCCTGAAG 763
954 AATTTGAAG 1013
764 GAACTGAAG 823
1014 TTGAGCATATTAACCCGAG 1073
824 CTTTACCTTTTAAACCCGAG 883
1074 CAGTGTGAG 1133
884 CAGGAG 943
1134 TGTAAACCCAG 1170
944 CATGAGTCTCACTGATCCCTGAG 1003
1171 CCCCATCAAG 1207
1004 ATCCAG 1063
1208 -----CTTATCCAG 1262

1064 TATCAGAGCTCCGAG 1123
1263 CATATGAG 1322
1124 CATATGAG 1183
1323 CAAGATGAG 1382
1184 CAAGATGAG 1243
1383 ACCGAG 1442
1244 GCGAG 1303
1443 CTTTAAACCCAG 1502
1304 CTTTGTCCAG 1363
1503 TGGCCGAG 1562
1364 TGGCCGAG 1423
1563 AGAAGTCAATCGTCTGAG 1622
1424 TGAAGTGTATGATCCAG 1483
1623 TATGAGGAG 1682
1484 CATCGAG 1543
1683 GGTGAAG 1742
1544 GGTGAAG 1603
1743 CTCCC 1747
1604 AGCCC 1608

Search completed: July 24, 2004, 16:48:44
Job time : 248.862 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2004, 13:04:51 ; Search time 1413.28 Seconds
(without alignments)
11355.361 Million cell updates/sec

Title: US-09-270-437d-8

Perfect score: 3283
Sequence: 1 ggcagcgagagcgagcgagga.....aaccttgaaatgttattc 3283

Scoring table:
IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3216467 seqs, 244419694 residues

Total number of hits satisfying chosen parameters: 6432934

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/prodata/2/pubpna/US05_NEW_PUB.seq:*
4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3282	100.0	3283	9	US-09-899-651-8
2	3255.4	99.2	3905	10	US-09-814-353-20241
3	3253.8	99.1	3667	15	US-10-097-340-146
4	3253.8	99.1	3667	17	US-10-648-593-48
5	3143	95.7	3412	9	US-09-899-651-6
6	3052.6	93.0	3684	9	US-09-764-864-329
7	1989.8	60.6	2250	9	US-09-764-864-332
8	1561.2	47.6	1707	16	US-10-262-445-39
9	1096.8	33.4	1186	9	US-09-764-864-749
10	697.2	21.2	1740	9	US-09-735-705-347
11	697.2	21.2	1740	9	US-09-850-716A-347
12	697.2	21.2	1740	9	US-09-897-778-347
13	697.2	21.2	1740	13	US-10-007-700-347
14	697.2	21.2	1740	15	US-10-117-982-347

15	697.2	21.2	1740	15	US-10-117-982-478	Sequence 478, App
16	697.2	21.2	1740	16	US-10-313-986-347	Sequence 347, App
17	697.2	21.2	1740	16	US-10-313-986-478	Sequence 478, App
18	697.2	21.2	1743	9	US-09-897-778-447	Sequence 447, App
19	697.2	21.2	1743	9	US-09-897-778-450	Sequence 450, App
20	697.2	21.2	1743	13	US-10-007-700-447	Sequence 447, App
21	697.2	21.2	1743	13	US-10-007-700-450	Sequence 450, App
22	697.2	21.2	1743	15	US-10-117-982-447	Sequence 447, App
23	697.2	21.2	1743	15	US-10-117-982-450	Sequence 450, App
24	697.2	21.2	1743	16	US-10-313-986-447	Sequence 447, App
25	697.2	21.2	1743	16	US-10-313-986-450	Sequence 450, App
26	697.2	21.2	1799	16	US-10-313-986-485	Sequence 485, App
27	697.2	21.2	34555	15	US-10-117-982-479	Sequence 479, App
28	697.2	21.2	34555	16	US-10-313-986-479	Sequence 479, App
29	697	21.2	1740	16	US-10-117-982-483	Sequence 483, App
30	697	21.2	1459	9	US-09-899-651-4	Sequence 4, App1
31	695.6	21.2	4181	9	US-09-735-705-175	Sequence 175, App
32	695.6	21.2	4181	9	US-09-954-456-715	Sequence 715, App
33	695.6	21.2	4181	9	US-09-850-716A-175	Sequence 175, App
34	695.6	21.2	4181	9	US-09-897-778-175	Sequence 175, App
35	695.6	21.2	4181	10	US-09-466-396A-175	Sequence 175, App
36	695.6	21.2	4181	10	US-10-007-700-175	Sequence 175, App
37	695.6	21.2	4181	13	US-10-117-982-175	Sequence 175, App
38	695.6	21.2	4181	15	US-10-313-986-175	Sequence 175, App
39	695.6	21.2	4181	16	US-10-116-802-145	Sequence 145, App
40	695.6	21.2	4181	13	US-09-850-716A-428	Sequence 428, App
41	694.2	21.1	1764	9	US-09-897-778-428	Sequence 428, App
42	694.2	21.1	1764	13	US-10-007-700-428	Sequence 428, App
43	694.2	21.1	1764	15	US-10-117-982-428	Sequence 428, App
44	694.2	21.1	1764	15	US-10-117-982-428	Sequence 428, App
45	694.2	21.1	1764	16	US-10-313-986-428	Sequence 428, App

ALIGNMENTS

RESULT 1
US-09-899-651-8
Sequence 8, Application US/09899651
Patent No. US2002011470A1
GENERAL INFORMATION:
APPLICANT: Chen, Yao-Tsang
APPLICANT: Tsang, Solam
APPLICANT: Stockert, Elisabeth
APPLICANT: Jager, Elke
APPLICANT: Knuth, Alexander
APPLICANT: Old, Lloyd J.
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
TITLE OF INVENTION: Associated
FILE REFERENCE: LUD 5538
CURRENT APPLICATION NUMBER: US/09/899,651
CURRENT FILING DATE: 2001-07-06
PRIORITY APPLICATION NUMBER: US/09/061,709
PRIORITY FILING DATE: 1998-04-17
NUMBER OF SEQ ID NOS: 8
SEQ ID NO 8
LENGTH: 3283
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
US-09-899-651-8

Query Match 100.0%: Score 3282; DB 9; Length 3283;
Best Local Similarity 100.0%: Pred No. 0;
Matches 3283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCAGCGAGGAGCGAGGAGCGCGGATACCGGCGGAGCGCGGCGCTCTGGGG 60
DB 1 GGCAGCGAGGAGCGAGGAGCGCGGATACCGGCGGAGCGCGGCGCTCTGGGG 60
QY 61 AAGAGCGAGTATGACAAAGCTTTACATCGGGAACCTAGGCCCGCGCTACCGCGAC 120

Db 61 AAGAGAGGATGATGAACAGCTTTTACATCGGGAACCTGAGCCCGCGCTACCGCCGAC 120
Qy 121 GACCTCGGCGAGCTCTTTGGGGAAGAAGCTGCCCTTGCGGGACAGCTCTGCTGAAG 180
Db 121 GACCTCGGCGAGCTCTTTGGGGAAGAAGCTGCCCTTGCGGGACAGCTCTGCTGAAG 180
Qy 181 TCCGGCTAAGCCTTGTGTGACTACCCGACAGAACTGGGCGCATCCGGGCACTGAGACC 240
Db 181 TCCGGCTAAGCCTTGTGTGACTACCCGACAGAACTGGGCGCATCCGGGCACTGAGACC 240
Qy 241 CTCTGGCTAAGTGAATGTCATGGGAAATCATGGAAGTGTATCTAGTCTCTAAA 300
Db 241 CTCTGGCTAAGTGAATGTCATGGGAAATCATGGAAGTGTATCTAGTCTCTAAA 300
Qy 301 AAGCTAAGAGAGAGAAAATTCAGATTCGAAACATCCCTCTCACTGACGTGGAGGTG 360
Db 301 AAGCTAAGAGAGAGAAAATTCAGATTCGAAACATCCCTCTCACTGACGTGGAGGTG 360
Qy 361 TTGATGATCTTTTGGCTCAATATGGGACAGTGGGAATGTGGAAACAAGTCAACAGAC 420
Db 361 TTGATGATCTTTTGGCTCAATATGGGACAGTGGGAATGTGGAAACAAGTCAACAGAC 420
Qy 421 TTGATGATCTTTTGGCTCAATATGGGACAGTGGGAATGTGGAAACAAGTCAACAGAC 420
Db 421 TTGATGATCTTTTGGCTCAATATGGGACAGTGGGAATGTGGAAACAAGTCAACAGAC 420
Qy 481 GAGAAAGCTAAGCGGCGATCAGTTTGAGAACTAATCTCTTCAAGATTTCTCAATCCCGAT 540
Db 481 GAGAAAGCTAAGCGGCGATCAGTTTGAGAACTAATCTCTTCAAGATTTCTCAATCCCGAT 540
Qy 541 GAAAGAGTGAAGTCCCTCGCCCTCGCCGACGAGGAGCCAGCTGGGGACCACTCTCCCG 600
Db 541 GAAAGAGTGAAGTCCCTCGCCCTCGCCGACGAGGAGCCAGCTGGGGACCACTCTCCCG 600
Qy 601 GAGCAAGGCGACGCGCTCGGCGCACTTCTCAAGCGCAGACAGATTGATTTCCGCTCGCG 660
Db 601 GAGCAAGGCGACGCGCTCGGCGCACTTCTCAAGCGCAGACAGATTGATTTCCGCTCGCG 660
Qy 661 ATCTGTGTCCTCCACCCAGTTTGTGTGTCATCATGGAAGAGGCGCTTGACATTAAG 720
Db 661 ATCTGTGTCCTCCACCCAGTTTGTGTGTCATCATGGAAGAGGCGCTTGACATTAAG 720
Qy 721 AACATACCTAAGAGAGCCAGTCCGCGGTAGATATCATAGAAAAAGAACTCTGAGCT 780
Db 721 AACATACCTAAGAGAGCCAGTCCGCGGTAGATATCATAGAAAAAGAACTCTGAGCT 780
Qy 781 GCAGAGAAGCTGTCAACATTCATGCAACCCGAGAGGGACTTTCTGAAGCATGCGCATG 840
Db 781 GCAGAGAAGCTGTCAACATTCATGCAACCCGAGAGGGACTTTCTGAAGCATGCGCATG 840
Qy 841 ATTCTTGAATCATGAGAAAGAGGCGATGAGAACCCAACTAGCCGAAAGATTTCTCTG 900
Db 841 ATTCTTGAATCATGAGAAAGAGGCGATGAGAACCCAACTAGCCGAAAGATTTCTCTG 900
Qy 901 AAAATCTGGCAACAATGCTTGTGTGAGAACTGATTGAAAAAGAGCAGAAATTTG 960
Db 901 AAAATCTGGCAACAATGCTTGTGTGAGAACTGATTGAAAAAGAGCAGAAATTTG 960
Qy 961 AAGAAATGGAACATGAAACAGGAGCCAGATTAACAATCTCATCTTTGACGATTTGAGC 1020
Db 961 AAGAAATGGAACATGAAACAGGAGCCAGATTAACAATCTCATCTTTGACGATTTGAGC 1020
Qy 1021 ATATTAACCCCGGAAAGAACATCACTGTGAAGGGCAGTTGAGGCTGTGCGAGTGT 1080
Db 1021 ATATTAACCCCGGAAAGAACATCACTGTGAAGGGCAGTTGAGGCTGTGCGAGTGT 1080
Qy 1081 GAATAGAGATTATGAAGAAGCTGCGTGAAGGCTTTGAAAAATGATAGCTGCGTTAAC 1140
Db 1081 GAATAGAGATTATGAAGAAGCTGCGTGAAGGCTTTGAAAAATGATAGCTGCGTTAAC 1140
Qy 1141 ACCCACTCCGATACCTTCTCAGCTGTACCCCATCAGCAGTTTGCCGCTCCGCAT 1200
Db 1141 ACCCACTCCGATACCTTCTCAGCTGTACCCCATCAGCAGTTTGCCGCTCCGCAT 1200

Db 1141 ACCCACTCCGATACCTTCTCAGCTGTACCCCATCAGCAGTTTGCCGCTCCGCAT 1200
Qy 1201 CATCACTCTTATCCAGAGCAGAGATTGTGAATCTTTCATCCCAACCAGGCTGTGGC 1260
Db 1201 CATCACTCTTATCCAGAGCAGAGATTGTGAATCTTTCATCCCAACCAGGCTGTGGC 1260
Qy 1261 GCCATCATCGGAAAGAGGGGCAACATCAAAACAGCTGGGAGATTGCGCGAGCCTCT 1320
Db 1261 GCCATCATCGGAAAGAGGGGCAACATCAAAACAGCTGGGAGATTGCGCGAGCCTCT 1320
Qy 1321 ATCAAGATTGCGCGGAGAGGCGCAACGTCAGCGAAAGGATGTCATCATCAACCGG 1380
Db 1321 ATCAAGATTGCGCGGAGAGGCGCAACGTCAGCGAAAGGATGTCATCATCAACCGG 1380
Qy 1381 CCACCGAAGCCAGTTCAAGGCCAGGGAAGATCTTTGGAAACTGAAAGAGAAAAC 1440
Db 1381 CCACCGAAGCCAGTTCAAGGCCAGGGAAGATCTTTGGAAACTGAAAGAGAAAAC 1440
Qy 1441 TTCTTTAAACCCAAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1500
Db 1441 TTCTTTAAACCCAAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1500
Qy 1501 GCTGCGCGGATGATGCAAAAGGTGGCAAGACCGTGAACGAACCTGCAAACTTAAACAGT 1560
Db 1501 GCTGCGCGGATGATGCAAAAGGTGGCAAGACCGTGAACGAACCTGCAAACTTAAACAGT 1560
Qy 1561 GCAGAAATCATGCTGCTCTGTCACAAACGCCAGATGAATGAAGAAATGATGTCAGA 1620
Db 1561 GCAGAAATCATGCTGCTCTGTCACAAACGCCAGATGAATGAAGAAATGATGTCAGA 1620
Qy 1621 ATTATCGGCGACTTTCTTGTGTCAGAGTGAACAGCGGAAATGAGAAATTTGACAA 1680
Db 1621 ATTATCGGCGACTTTCTTGTGTCAGAGTGAACAGCGGAAATGAGAAATTTGACAA 1680
Qy 1681 CAGGTGAAGCAGCAGAGAGAGAAATACCTCAGGAGTGCCTCAGAGCGAGCAAGTGA 1740
Db 1681 CAGGTGAAGCAGCAGAGAGAGAAATACCTCAGGAGTGCCTCAGAGCGAGCAAGTGA 1740
Qy 1741 GGCCTCCACAGGACCCAGCAAAACAAACGATGATGATGATGATGATGATGATGAT 1800
Db 1741 GGCCTCCACAGGACCCAGCAAAACAAACGATGATGATGATGATGATGATGATGAT 1800
Qy 1801 TGAAGCAAAACGACGACCGAGATCGGAGCAAAACCAAAAGCATCTGAGAAATGAGAA 1860
Db 1801 TGAAGCAAAACGACGACCGAGATCGGAGCAAAACCAAAAGCATCTGAGAAATGAGAA 1860
Qy 1861 GTCTGCGAGGCGGCGAGGGAATCTGCGAGGCTTGAGAAACCCAGGCGCGAGAGG 1920
Db 1861 GTCTGCGAGGCGGCGAGGGAATCTGCGAGGCTTGAGAAACCCAGGCGCGAGAGG 1920
Qy 1921 GCGGGAAGGTCAAGCAGTTTGCAGAAACCAACGAGCCCTGCTCCGCCCCCAGGCG 1980
Db 1921 GCGGGAAGGTCAAGCAGTTTGCAGAAACCAACGAGCCCTGCTCCGCCCCCAGGCG 1980
Qy 1981 TTCTGAGGCTTCAAGCATTCATCAACATCCATCCGATCTCTCTGTAACCTCCAGCA 2040
Db 1981 TTCTGAGGCTTCAAGCATTCATCAACATCCATCCGATCTCTCTGTAACCTCCAGCA 2040
Qy 2041 CGCTATCCCTTTAGTTGAACATAATAGTGAACGTTTCAAAAGCCAGCAAAATGAC 2100
Db 2041 CGCTATCCCTTTAGTTGAACATAATAGTGAACGTTTCAAAAGCCAGCAAAATGAC 2100
Qy 2101 ACCCTTTTCTGTGGCAATGCTCTGTGACATGTGTGACATTAATTAAGAAATG 2160
Db 2101 ACCCTTTTCTGTGGCAATGCTCTGTGACATGTGTGACATTAATTAAGAAATG 2160
Qy 2161 TTAAGATATGTGCGCTGTGGTTACACAGGAGCGTCCAGCGGATATATATTAAGAAAT 2220
Db 2161 TTAAGATATGTGCGCTGTGGTTACACAGGAGCGTCCAGCGGATATATATTAAGAAAT 2220
Qy 2221 AATATATCAATTAATCAATTAATCAATTAATCAATTAATTAATTAATTAATTAAT 2280
Db 2221 AATATATCAATTAATCAATTAATCAATTAATCAATTAATTAATTAATTAATTAAT 2280

2281 TTAAAGAAAGAGGCTTTCTAGACTTTAAAGATAAAGCTTTGGAGGCTCAAG 2340
2281 TTAAAGAAAGAGGCTTTCTAGACTTTAAAGATAAAGCTTTGGAGGCTCAAG 2340
2341 TGTAGAGGAGCTTTGAGGCAACCCGCAAAATTCACCCAGAGGAAATCGTCGGA 2400
2341 TGTAGAGGAGCTTTGAGGCAACCCGCAAAATTCACCCAGAGGAAATCGTCGGA 2400
2401 AGGACATCAAGGAGCTTGTGATCACTGTGTATGATCAAGAGGATACCGTCTCT 2460
2401 AGGACATCAAGGAGCTTGTGATCACTGTGTATGATCAAGAGGATACCGTCTCT 2460
2461 TGAAGAGGAACTGTGATCACTGTGTATGATCAAGAGGATACCGTCTCT 2520
2461 TGAAGAGGAACTGTGATCACTGTGTATGATCAAGAGGATACCGTCTCT 2520
2521 TTCAAGGCTTTAACTGTGATCACTGTGTATGATCAAGAGGATACCGTCTCT 2580
2521 TTCAAGGCTTTAACTGTGATCACTGTGTATGATCAAGAGGATACCGTCTCT 2580
2581 TATCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2640
2581 TATCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2640
2641 CCTCCATCTCAATCCGCTATCTAGAGACCCCTCCCTCCCTCCCTCCCTCCCT 2700
2641 CCTCCATCTCAATCCGCTATCTAGAGACCCCTCCCTCCCTCCCTCCCTCCCT 2700
2701 GTATCAATCAACCAAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAA 2760
2701 GTATCAATCAACCAAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAA 2760
2761 GGTACTCAAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAA 2820
2761 GGTACTCAAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAA 2820
2821 GGAACCCAGCAACCAACCAATCAACCAACCAACCAACCAACCAACCAACCA 2880
2821 GGAACCCAGCAACCAACCAATCAACCAACCAACCAACCAACCAACCAACCA 2880
2881 ATTTTGTCTTTTGTGATTTTGTGATTTTGTGATTTTGTGATTTTGTGATTT 2940
2881 ATTTTGTCTTTTGTGATTTTGTGATTTTGTGATTTTGTGATTTTGTGATTT 2940
2941 TTTAAAAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGG 3000
2941 TTTAAAAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGG 3000
3001 AAATTCAGAGATTTTAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGG 3060
3001 AAATTCAGAGATTTTAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGG 3060
3061 CCTCAGCAGCTTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3120
3061 CCTCAGCAGCTTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3120
3121 TTTTATTTTAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAA 3180
3121 TTTTATTTTAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAA 3180
3181 GAGAGGATGACAGTCCAGAGTGTGCAAGAGTGTGCAAGAGTGTGCAAGAGT 3240
3181 GAGAGGATGACAGTCCAGAGTGTGCAAGAGTGTGCAAGAGTGTGCAAGAGT 3240
3241 CCNATATGAGCTTTCTTTGACAAACCTTGAAATGTTATTT 3283
3241 CCNATATGAGCTTTCTTTGACAAACCTTGAAATGTTATTT 3283
3241 CCNATATGAGCTTTCTTTGACAAACCTTGAAATGTTATTT 3283

RESULT 2
US-09-814-353-20241
; Sequence 20241, Application US/09814353

Publication No. US20030165831A1
GENERAL INFORMATION:
APPLICANT: Lee, John
APPLICANT: Thompson, Pamela
APPLICANT: Lillie, James
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
FILE REFERENCE: MRI-006B
CURRENT APPLICATION NUMBER: US/09/814,353
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/191,031
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: US 60/207,124
PRIOR FILING DATE: 2000-05-25
PRIOR APPLICATION NUMBER: US 60/211,940
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: US 60/216,820
PRIOR FILING DATE: 2000-07-07
PRIOR APPLICATION NUMBER: US 60/220,661
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: US 60/257,672
PRIOR FILING DATE: 2000-12-21
NUMBER OF SEQ ID NOS: 22037
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 20241
LENGTH: 3905
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 1, 3897, 3898, 3899, 3900, 3901, 3902, 3903, 3904, 3905
OTHER INFORMATION: n = A,T,C or G
US-09-814-353-20241
Query Match 99.2%; Score 3255.4; DB 10; Length 3905;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3278; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
3 CAGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 62
369 CAGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 428
63 GAGACGATGATTAACAAGCTTTTACATCGGAAACCTGAGCCCGCTGACCGCGA 122
429 GAGACGATGATTAACAAGCTTTTACATCGGAAACCTGAGCCCGCTGACCGCGA 488
123 CCGCGGAGCTCTTTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 182
489 CCGCGGAGCTCTTTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 548
183 CCGCTAGCCTTGTGAGTACCCGACGAGAGCTGAGGAGGAGGAGGAGGAGGAGG 242
549 CCGCTAGCCTTGTGAGTACCCGACGAGAGCTGAGGAGGAGGAGGAGGAGGAGG 608
243 CTGGGTAAGTGAATTCGATGGAAGATCATGAAGTTGATTACTCACTCTTAAAA 302
609 CTGGGTAAGTGAATTCGATGGAAGATCATGAAGTTGATTACTCACTCTTAAAA 668
303 GCTAAGAGAGGAAATTCAGATTGGAAGATTCCTCTCAGCTGAGGAGGAGGAGG 362
669 GCTAAGAGAGGAAATTCAGATTGGAAGATTCCTCTCAGCTGAGGAGGAGGAGG 728
363 GATGAGCTTTGGCTCAATATGAGGAGTGTGAGGAGTGTGAGGAGTGTGAGGAG 422
729 GATGAGCTTTGGCTCAATATGAGGAGTGTGAGGAGTGTGAGGAGTGTGAGGAG 788
423 AGAAACCGCGCTTGTACATATGCAATATGCAATATGCAATATGCAATATGCAAT 482
789 AGAAACCGCGCTTGTACATATGCAATATGCAATATGCAATATGCAATATGCAAT 848
483 GAGGCTAAGCGGAGATGTTGAGAGATCTCTTCAAGATTTCTTCAATCCGATGA 542

Db 849 GAAAGCTAGCGGGGATCAGTTTGAAATCTCCTTCAAGATTTCCTACATCCCGAGTGA 908
QY 543 AGAGGTAGCTCCCTTTCGCCCCCTCAGCGAGCCGACGCTGGGGACCACTTTCGCGGGA 602
Db 909 AGAGGTAGCTCCCTTTCGCCCCCTCAGCGAGCCGACGCTGGGGACCACTTTCGCGGGA 968
QY 603 GCAAGGCAAGCCCTTGGGGGCACTTCTCAGGCGCAGATGATTTCCCGTGGCGAT 662
Db 969 GCAAGGCAAGCCCTTGGGGGCACTTCTCAGGCGCAGATGATTTCCCGTGGCGAT 1028
QY 663 CTTGATCCCGCAGCTTGTGTGTGTCATCATCGAAAGAGGAGCTTGACATTAAGAA 722
Db 1029 CTTGATCCCGCAGCTTGTGTGTGTCATCATCGAAAGAGGAGCTTGACATTAAGAA 1088
QY 723 CATCATTAAGCAGACCCAGTCCCGGGTATCATTCATTAAGAAAGAACTTGGAGCTGC 782
Db 1089 CATCATTAAGCAGACCCAGTCCCGGGTATCATTCATTAAGAAAGAACTTGGAGCTGC 1148
QY 783 AGAGAGGCTGTGACCATCCATCCAGTCCAGAGGGGAGCTTCTGAGCATCCGATGAT 842
Db 1149 AGAGAGGCTGTGACCATCCATCCAGTCCAGAGGGGAGCTTCTGAGCATCCGATGAT 1208
QY 843 TCTTGAATCATGCGAAGAGGAGGAGATGAGACCAAACTAGCCGAGAGATTCCTTGA 902
Db 1209 TCTTGAATCATGCGAAGAGGAGGAGATGAGACCAAACTAGCCGAGAGATTCCTTGA 1268
QY 903 AATCTTGGCACAATGCGCTTGTGTGTGAAAGCTGATTTGAAAGAAAGGACGAAATTTGA 962
Db 1269 AATCTTGGCACAATGCGCTTGTGTGTGAAAGCTGATTTGAAAGAAAGGACGAAATTTGA 1328
QY 963 GAAATTTGAACATGAAAGAGGAGGAGCAAGATTAACAATCTCATCTTTCAGGATTTGAGAT 1022
Db 1329 GAAATTTGAACATGAAAGAGGAGGAGCAAGATTAACAATCTCATCTTTCAGGATTTGAGAT 1388
QY 1023 ATACAAACCGGAAAGAACCATCATCTGTGAAGGAGCAAGTTGAGGCTGTGCCAGTGTGA 1082
Db 1389 ATACAAACCGGAAAGAACCATCATCTGTGAAGGAGCAAGTTGAGGCTGTGCCAGTGTGA 1448
QY 1083 GATAGAGATTTAGAGAGCTGCTGTGAGGCTTTGAAATGATTTGCTGGCTGTTAAAGC 1142
Db 1449 GATAGAGATTTAGAGAGCTGCTGTGAGGCTTTGAAATGATTTGCTGGCTGTTAAAGC 1508
QY 1143 GCACCTCCGAGTACTTCTCCAGCTGTACCCCATCAGCAGTTTGGCCGCTTCCCGATCA 1202
Db 1509 GCACCTCCGAGTACTTCTCCAGCTGTACCCCATCAGCAGTTTGGCCGCTTCCCGATCA 1568
QY 1203 TCACTCTTATCCAGAGCAGAGATTTGTGATCTTTCATCCAAACCCAGGCTGTGGGCGC 1262
Db 1569 TCACTCTTATCCAGAGCAGAGATTTGTGATCTTTCATCCAAACCCAGGCTGTGGGCGC 1628
QY 1263 CATCATTCGAGAGAGGAGGAGCACAATCAAAAGCTGGGAGATTTGGCCGAGACCTTAT 1322
Db 1629 CATCATTCGAGAGAGGAGGAGCACAATCAAAAGCTGGGAGATTTGGCCGAGACCTTAT 1688
QY 1323 CAAGATTTGCCCTCGGAGAGGAGCCAGAGCTCAGCAGAAAGGATGTCATCATCAACCGGCGC 1382
Db 1689 CAAGATTTGCCCTCGGAGAGGAGCCAGAGCTCAGCAGAAAGGATGTCATCATCAACCGGCGC 1748
QY 1383 ACCGGAAGCCAGATTCAAGGCCAGAGGAGCGGATCTTTGGGAAACTGAAAGAGAAAATT 1442
Db 1749 ACCGGAAGCCAGATTCAAGGCCAGAGGAGCGGATCTTTGGGAAACTGAAAGAGAAAATT 1808
QY 1443 CTTTAAACCCAAAGAGAGATGAAAGCTGGAGAGGCAATTCAGATGCGCTTTCACAGC 1502
Db 1809 CTTTAAACCCAAAGAGAGATGAAAGCTGGAGAGGCAATTCAGATGCGCTTTCACAGC 1868
QY 1503 TGGCCGGGTGATTTGGCAAAGGTGCAAGAGCGTGAAGCAACTGCAAGACTTAAACCACTGC 1562
Db 1869 TGGCCGGGTGATTTGGCAAAGGTGCAAGAGCGTGAAGCAACTGCAAGACTTAAACCACTGC 1928
QY 1563 AGAAGTCATGTCGTCGTGTGACCAAGCCAGATGAAATGAGGAAGTGTGTGAGAT 1622
Db 1929 AGAAGTCATGTCGTCGTGTGACCAAGCCAGATGAAATGAGGAAGTGTGTGAGAT 1988

QY 1623 TATCGGGCACTTCTTTCCTAGCCAGACTGACAGCGCAAGATCAGGGAAATTTGACACA 1682
Db 1989 TATCGGGCACTTCTTTCCTAGCCAGACTGACAGCGCAAGATCAGGGAAATTTGACACA 2048
QY 1683 GGTGAAGCAGCAGAGAGCAAAATTCCTTCAGGGAGTGGCTTCAAGCGCAGCAAGTGAAG 1742
Db 2049 GGTGAAGCAGCAGAGAGCAAAATTCCTTCAGGGAGTGGCTTCAAGCGCAGCAAGTGAAG 2108
QY 1743 CTCCCAAGGCAACAGCAAAACAACGATGATTAAGCTTTCACACCTGACGAATG 1802
Db 2109 CTCCCAAGGCAACAGCAAAACAACGATGATTAAGCTTTCACACCTGACGAATG 2168
QY 1803 AGACCAACGCAAGCAGCAGATGAGAGCAACCAAGACCAATCTGAGAAATGAAAGT 1862
Db 2169 AGACCAACGCAAGCAGCAGATGAGAGCAACCAAGACCAATCTGAGAAATGAAAGT 2228
QY 1863 CTGCGGAGCGGCGCAGAGGACCTGCGGAGGCGCTGAGAACCCAGAGGCGCAGAGGAGC 1922
Db 2229 CTGCGGAGCGGCGCAGAGGACCTGCGGAGGCGCTGAGAACCCAGAGGCGCAGAGGAGC 2288
QY 1923 GGGGAAGGTCAAGCAGTTTGCAGAACCAACGAGCCCGCTCCGCCCCCAGGCTT 1982
Db 2289 GGGGAAGGTCAAGCAGTTTGCAGAACCAACGAGCCCGCTCCGCCCCCAGGCTT 2348
QY 1983 CTGAGGCTTCAAGCCATCCATTCACATCCATCTGGAATCTTCTGAACTCCCAAGAG 2042
Db 2349 CTGAGGCTTCAAGCCATCCATTCACATCCATCTGGAATCTTCTGAACTCCCAAGAG 2408
QY 2043 CTATCCCTTTTATGTTGAATCAATAGGTGAACGTGTTCAAGCAAGCAAGCAATGACAC 2102
Db 2409 CTATCCCTTTTATGTTGAATCAATAGGTGAACGTGTTCAAGCAAGCAAGCAATGACAC 2468
QY 2103 CTTTTCCTGAGCAAAATGCTCTGTACATGTGTATCATTTAGAAAGGAGATGTT 2162
Db 2469 CTTTTCCTGAGCAAAATGCTCTGTACATGTGTATCATTTAGAAAGGAGATGTT 2528
QY 2163 AAGATATGTGCTGTGGGTTTCAAGGGTGCCTGAGGGGTATATTTAGAAATTA 2222
Db 2529 AAGATATGTGCTGTGGGTTTCAAGGGTGCCTGAGGGGTATATTTAGAAATTA 2588
QY 2223 TATATCAAAATTAATCACTCAATCTCAATTTTAAATTTTAAATTTTCTTTT 2282
Db 2589 TATATCAAAATTAATCACTCAATCTCAATTTTAAATTTTAAATTTTCTTTT 2648
QY 2283 AAAGAGAAAGCAGGCTTTCAGACTTAAAGATTAAGATCTTTGGAGGCTCAAGGTTG 2342
Db 2649 AAAGAGAAAGCAGGCTTTCAGACTTAAAGATTAAGATCTTTGGAGGCTCAAGGTTG 2708
QY 2343 TAGAGAGAGCTTTGAGGCGACCCGCAAAATTCACCCAGAGGAAATCTGTCGGAAG 2402
Db 2709 TAGAGAGAGCTTTGAGGCGACCCGCAAAATTCACCCAGAGGAAATCTGTCGGAAG 2768
QY 2403 GACACCTCAGGCAAGTCTGGATCACTGTGTATGTCAAGAGAGGATACCGTCTCTTG 2462
Db 2769 GACACCTCAGGCAAGTCTGGATCACTGTGTATGTCAAGAGAGGATACCGTCTCTTG 2828
QY 2463 AAGAGAAAATCTGTCACTCTCATGCTGTAGCTCATACCAATTTCTCTTGCTT 2522
Db 2829 AAGAGAAAATCTGTCACTCTCATGCTGTAGCTCATACCAATTTCTCTTGCTT 2888
QY 2523 CACAGGTTTAAACTGTTTGTGATCTGCTATCTGCTATCTGCTCTGCTTGA 2582
Db 2889 CACAGGTTTAAACTGTTTGTGATCTGCTATCTGCTATCTGCTCTGCTTGA 2948
QY 2583 TCTCTCCCTCCCTCCCTCCCTCCCTCTTCTCAATCTCAATCTTGAATTTCTCTATCC 2642
Db 2949 TCTCTCCCTCCCTCCCTCCCTCCCTCTTCTCAATCTCAATCTTGAATTTCTCTATCC 3008
QY 2643 CTCCATCTCAATCCGATATCAGACCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2702
Db 3009 CTCCATCTCAATCCGATATCAGCA-CCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 3067

QY 2703 ATCATATACACAAAGGAAACAAAGCGAAACACAAACAGCTCAACTTACCTTGG 2762
Db 3068 ATCATATACACAAAGGAAACAAAGCGAAACACAAACAGCTCAACTTACCTTGG 3127
QY 2763 TTACTCAGAAAGCAAGGTCATGTAATCTGTCTTACGCTTTGGAGAGGAAAGAG 2822
Db 3128 TTACTCAGAAAGCAAGGTCATGTAATCTGTCTTACGCTTTGGAGAGGAAAGAG 3187
QY 2823 AACCCAGCAACCAACCAATCAACCAACCAAGAAATTTCCCAATGAAAGATGAT 2882
Db 3188 AACCCAGCAACCAACCAATCAACCAACCAAGAAATTTCCCAATGAAAGATGAT 3247
QY 2883 TTGTCTTTTTCGATTTTGTGTATTAAGCCATCAATATTTAGCAAAATGATCTTCT 2942
Db 3248 TTGTCTTTTTCGATTTTGTGTATTAAGCCATCAATATTTAGCAAAATGATCTTCT 3307
QY 2943 T-AAAAAAAAAATGTGAGGAAAGTAAATTTACCAAGGTTGTGGCCAGGCGCTTA 3001
Db 3308 TAAAAAAAAAATGTGAGGAAAGTAAATTTACCAAGGTTGTGGCCAGGCGCTTA 3367
QY 3002 AATTCAGAGATTTTAAACGAGAAACACACAGAAAGAGCTTCTCAGGTGTTTTAC 3061
Db 3368 AATTCAGAGATTTTAAACGAGAAACACACAGAAAGAGCTTCTCAGGTGTTTTAC 3427
QY 3062 CTCAGACCTTGTCTTGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3121
Db 3428 CTCAGACCTTGTCTTGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3487
QY 3122 TTTTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3181
Db 3488 TTTTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3547
QY 3182 AGAAGGTGACAGTCCAGTGTGCAAGCTGTCTGTAATGTCTTCCGCTTACAGAGAT 3241
Db 3548 AGAAGGTGACAGTCCAGTGTGCAAGCTGTCTGTAATGTCTTCCGCTTACAGAGAT 3607
QY 3242 CNAATATGCGCTTCTTTTGAACAAACCTTGAATGTTATTT 3283
Db 3608 CNAATATGCGCTTCTTTTGAACAAACCTTGAATGTTATTT 3649

RESULT 3
US-10-097-340-146
; Publication 146, Application US/10097340
; GENERAL INFORMATION:
; APPLICANT: JOMI MONAHAN
; APPLICANT: Manjula GANNAVAPU
; APPLICANT: Sebastian HOERSCHE
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISSEY
; APPLICANT: Peter OLANDT
; APPLICANT: Aml SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14

; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 146
; LENGTH: 3667
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-097-340-146

Query Match 99.1%; Score 3253.8; DB 15; Length 3667;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 3277; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 3 CAGCGAGAGAGCGAGAGAGCGCGGTACCGCGCGCGAGAGCGCGCGCGCTCTCGGGGAA 62
Db 369 CAGCGAGAGAGCGAGAGAGCGCGGTACCGCGCGCGAGAGCGCGCGCGCTCTCGGGAA 428
QY 63 GAGACGATGATGAACAAGCTTTTACATCGGAACTTGAAGCCCGCGGTCAACCGCGAGA 122
Db 429 GAGACGATGATGAACAAGCTTTTACATCGGAACTTGAAGCCCGCGGTCAACCGCGAGA 488
QY 123 CTTCCCGGACCTCTTGGGGGACAGAAAGTCCCTTGGCGGAGACAGTCTTCTGAAAGTC 182
Db 489 CTTCCCGGACCTCTTGGGGGACAGAAAGTCCCTTGGCGGAGACAGTCTTCTGAAAGTC 548
QY 183 CCGCTACGCTTGTGACTACCCCGACAGAACTGGGCGATCCGCGCATCGAGACCT 242
Db 549 CCGCTACGCTTGTGACTACCCCGACAGAACTGGGCGATCCGCGCATCGAGACCT 608
QY 243 CTGGGTAAGGTGAATGATGGAATCATGAAAGTTGATTTCTCACTTCTTAAAAA 302
Db 609 CTGGGTAAGGTGAATGATGGAATCATGAAAGTTGATTTCTCACTTCTTAAAAA 668
QY 303 GCTAAGAGAGGAAATTCAGATTCGAAACATCCCTCTCACTGTGAGTGGAGGTGTT 362
Db 669 GCTAAGAGAGGAAATTCAGATTCGAAACATCCCTCTCACTGTGAGTGGAGGTGTT 728
QY 363 GGATGACCTTTTGGCTCAATATGAGCAAGTGAAGATGAGCAAGTCAACAGACAC 422
Db 729 GGATGACCTTTTGGCTCAATATGAGCAAGTGAAGATGAGCAAGTCAACAGACAC 788
QY 423 AGAAACCGCGCTTGTCAAGTCAATATGCAACAAGAAAGAAATTAAGCATGGA 482
Db 789 AGAAACCGCGCTTGTCAAGTCAATATGCAACAAGAAAGAAATTAAGCATGGA 848
QY 483 GAAAGTAAAGGGGATCAGTTGAGAACTACTCTTCAAGATTTCTTCAATCCCGATGA 542
Db 849 GAAAGTAAAGGGGATCAGTTGAGAACTACTCTTCAAGATTTCTTCAATCCCGATGA 908
QY 543 AGAGTGAAGTCCCTTTCGCGCGCTTCAAGAGAGCGAGCGTGGGACCACTTCCGGGA 602
Db 909 AGAGTGAAGTCCCTTTCGCGCGCTTCAAGAGAGCGAGCGTGGGACCACTTCCGGGA 968
QY 603 GCAAGGCGACGCGCTTGGGGGACCTTTCAGGCGACAGATGATTTCCCGTGGGAT 662
Db 969 GCAAGGCGACGCGCTTGGGGGACCTTTCAGGCGACAGATGATTTCCCGTGGGAT 1028
QY 663 CCGGTCCCGACCGAGTTTGTGTCCTCACTCGAAAGAGAGGCTTGAACCTAAGAA 722
Db 1029 CCGGTCCCGACCGAGTTTGTGTCCTCACTCGAAAGAGAGGCTTGAACCTAAGAA 1088
QY 723 CATCATAGACAGACCCAGTCCCGGGTATATTCATAGAAAGAACTCTGAGACTGC 782
Db 1089 CATCATAGACAGACCCAGTCCCGGGTATATTCATAGAAAGAACTCTGAGACTGC 1148
QY 783 AGAAGAGCTGTACCATTCATGCCACCCAGAGGGGACTTCTGAAGCATGCGCATGAT 842

```
Db 1149 AGAAGACCTGTCACTCCATGCGACCCAGAGGGACCTTTGAAAGCATGCGCATGAT 1208
QY 843 TCTTGAATCATGACGAAAGAGGAGATGAGACCAAACTAGCCGAGAGATTCCTGTGA 902
Db 1209 TCTTGAATCATGACGAAAGAGGAGATGAGACCAAACTAGCCGAGAGATTCCTGTGA 1268
QY 903 AATCTGGACACAAATGGCTTGGTGGAGACTGATTTGGAAAAGAGGACAAATTTGAA 962
Db 1269 AATCTGGACACAAATGGCTTGGTGGAGACTGATTTGGAAAAGAGGACAAATTTGAA 1328
QY 963 GAAATTTGAACATGAAACAGGGACCAAGATTAACAATCTCATCTTTGACAGATTTGAGCAT 1022
Db 1329 GAAATTTGAACATGAAACAGGGACCAAGATTAACAATCTCATCTTTGACAGATTTGAGCAT 1388
QY 1023 ATACAAACCCGAAAGAAACATCACTGTGAAAGGACAGTTGAGGCTGTGCAAGTCTGA 1082
Db 1389 ATACAAACCCGAAAGAAACATCACTGTGAAAGGACAGTTGAGGCTGTGCAAGTCTGA 1448
QY 1083 GATGAGATTTGAAAGAGCTGCGTGAGAGCTTTGAAAATGATATGCTGGCTGTTAACAC 1142
Db 1449 GATGAGATTTGAAAGAGCTGCGTGAGAGCTTTGAAAATGATATGCTGGCTGTTAACAC 1508
QY 1143 CCACCTCCGAGATCTCTCCAGCCTGTACCCCATCACCAGTTTGGCCCGTCCCGCATCA 1202
Db 1509 CCACCTCCGAGATCTCTCCAGCCTGTACCCCATCACCAGTTTGGCCCGTCCCGCATCA 1568
QY 1203 TCACCTTATCCAGACAGAGATTTGAAATCTTTTATCCCAACCCAGGCTGTGGGGC 1262
Db 1569 TCACCTTATCCAGACAGAGATTTGAAATCTTTTATCCCAACCCAGGCTGTGGGGC 1628
QY 1263 CATCATCCGGAAGAAAGGGGACACATCAACAGCTGGGAGATTCGCGGAGCCTCTAT 1322
Db 1629 CATCATCCGGAAGAAAGGGGACACATCAACAGCTGGGAGATTCGCGGAGCCTCTAT 1688
QY 1323 CAAGATGGCCCTGCGGAAAGGCCAGACGTCAAGCAAAAGATGTCATCATCCGAGGC 1382
Db 1689 CAAGATGGCCCTGCGGAAAGGCCAGACGTCAAGCAAAAGATGTCATCATCCGAGGC 1748
QY 1383 ACCGGAAGCCCAATTCAGAGCCCAAGGACGATCTTTGGGAACTGAAAGAGAAACTT 1442
Db 1749 ACCGGAAGCCCAATTCAGAGCCCAAGGACGATCTTTGGGAACTGAAAGAGAAACTT 1808
QY 1443 CTTTAAACCCCAAGAAAGATGTAAGCTGGAAGGCAATCAGAGTCCCTCTTCCACAGC 1502
Db 1809 CTTTAAACCCCAAGAAAGATGTAAGCTGGAAGGCAATCAGAGTCCCTCTTCCACAGC 1868
QY 1503 TGGCCGGGATTTGGCAAGGTGGCAAGACCGTGAACGAATGACAGACTTAAACAAGTGC 1562
Db 1869 TGGCCGGGATTTGGCAAGGTGGCAAGACCGTGAACGAATGACAGACTTAAACAAGTGC 1928
QY 1563 AGAAGTCATGTCCTGTCGACCAAAAGCCCAATGAAATGAGAAAGTATGCTCAGAAAT 1622
Db 1929 AGAAGTCATGTCCTGTCGACCAAAAGCCCAATGAAATGAGAAAGTATGCTCAGAAAT 1988
QY 1623 TATCGGACCTTCTTGTAGCCAGACTGACAGCCCAAGATCAGGGAATTTGTACACA 1682
Db 1989 TATCGGACCTTCTTGTAGCCAGACTGACAGCCCAAGATCAGGGAATTTGTACACA 2048
QY 1683 GGTGAGACGACGAGACAGAAATACCTCAGGAGTGCCTCAACGCGACGACAAAGTGAAG 1742
Db 2049 GGTGAGACGAGACGAGAAATACCTCAGGAGTGCCTCAACGCGACGACAAAGTGAAG 2108
QY 1743 CTCCTCAGGACACAGCAAAACACAGATGATGATGAGCTTTCACAACTGACAGATG 1802
Db 2109 CTCCTCAGGACACAGCAAAACACAGATGATGATGAGCTTTCACAACTGACAGATG 2168
QY 1803 AGACCAAAACGACACGACGACATCGGAGGACAAACCAAGACCAATCTGAGAAATGAGAGT 1862
Db 2169 AGACCAAAACGACACGACGACATCGGAGGACAAACCAAGACCAATCTGAGAAATGAGAGT 2228
QY 1863 CTCGAGGAGGACGACGAGACTTGCAGAGCCCTGAGAACCCCAAGGAGCGAGAGGAGC 1922
Db 2229 CTCGAGGAGGACGACGAGACTTGCAGAGCCCTGAGAACCCCAAGGAGCGAGAGGAGC 2288
QY 1923 GGGGAGGTCAGGACGAGTTTGGCAGAACCAACGAGCCCGGCTCCGCCCCAGGCTT 1982
Db 2289 GGGGAGGTCAGGACGAGTTTGGCAGAACCAACGAGCCCGGCTCCGCCCCAGGAGCTT 2348
QY 1983 CTGAGGCTTACGACATCCACTTCAACATGCACTGCGATCTCTCTGAACTCCCAACAGC 2042
Db 2349 CTGAGGCTTACGACATCCACTTCAACATGCACTGCGATCTCTCTGAACTCCCAACAGC 2408
QY 2043 CTATCCCTTTAGTTGAACATACATAGGTGAACGTGTTCAAGCCAAACAAATGACAC 2102
Db 2409 CTATCCCTTTAGTTGAACATACATAGGTGAACGTGTTCAAGCCAAACAAATGACAC 2468
QY 2103 CCTTTTCTGTGGCAATCGTCTGTGACATGTGTGACATATTAATTTTCTTTT 2162
Db 2469 CCTTTTCTGTGGCAATCGTCTGTGACATGTGTGACATATTAATTTTCTTTT 2528
QY 2163 AAGATATGTGCTGTGGTTACACAGGTCCTGACAGCGGTAAATATTTAGAAATTA 2222
Db 2529 AAGATATGTGCTGTGGTTACACAGGTCCTGACAGCGGTAAATATTTAGAAATTA 2588
QY 2223 TATATCAATATACCTCACTAACTCCAAATTTTAAATCAATTAATTTTCTTTT 2282
Db 2589 TATATCAATATACCTCACTAACTCCAAATTTTAAATCAATTAATTTTCTTTT 2648
QY 2283 AAAGGAAAGAGGCTTTTCTGACATTTAAAGATTAAGTCTTGGAGGTCTACGCTG 2342
Db 2649 AAAGGAAAGAGGCTTTTCTGACATTTAAAGATTAAGTCTTGGAGGTCTACGCTG 2708
QY 2343 TAGAGAGAGCTTTAGAGCCACCCGCACAAAATTCACCAAGAGGAAATCTGCGGAG 2402
Db 2709 TAGAGAGAGCTTTAGAGCCACCCGCACAAAATTCACCAAGAGGAAATCTGCGGAG 2768
QY 2403 GACACTCACGAGATTTGAGATCACTGTGATGTCAACAGAGGAAATCCGTCTCTTG 2462
Db 2769 GACACTCACGAGATTTGAGATCACTGTGATGTCAACAGAGGAAATCCGTCTCTTG 2828
QY 2463 AAGAGAAATCTGTCACTCTCATGCTGTGTAGCTCATACCATATTTCTTTGCTT 2522
Db 2829 AAGAGAAATCTGTCACTCTCATGCTGTGTAGCTCATACCATATTTCTTTGCTT 2888
QY 2523 CACAGTTTAACTGTGTTTTTGAATGCTGATATATATCTGCTCTCTGTTA 2582
Db 2889 CACAGTTTAACTGTGTTTTTGAATGCTGATATATATCTGCTCTCTGTTA 2948
QY 2583 TCTCTCCCTCCCTCCCTCCCTCTCTCTCATCTCAATCTTTGAAATTTCTCATCC 2642
Db 2949 TCTCTCCCTCCCTCCCTCCCTCTCTCTCATCTCAATCTTTGAAATTTCTCATCC 3008
QY 2643 CTCATCTCAATCCGATATCTAGCAACCCCGGAGGAGGAGGAGGAGGAGGAGGAGG 2702
Db 3009 CTCATCTCAATCCGATATCTAGCAACCCCGGAGGAGGAGGAGGAGGAGGAGGAGG 3067
QY 2703 ATCACTTCAACAAAGAAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2762
Db 3068 ATCACTTCAACAAAGAAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3127
QY 2763 TTAATCAAAAGAACAGAGTCAATGCTTGTCTAGGCTTTTGAAGAGGAAACAG 2822
Db 3128 TTAATCAAAAGAACAGAGTCAATGCTTGTCTAGGCTTTTGAAGAGGAAACAG 3187
QY 2823 AACCCACAAACCAACCAATCAACCAACCAACCAACCAACCAACCAATTCACATGAAAGATGAT 2882
Db 3188 AACCCACAAACCAACCAATCAACCAACCAACCAACCAACCAACCAATTCACATGAAAGATGAT 3247
QY 2883 TTTGTCTTTTGAATTTGTGTATAGCATGATATTCAGCAAAATGATCTTCTT 2942
Db 3248 TTTGTCTTTTGAATTTGTGTATAGCATGATATTCAGCAAAATGATCTTCTT 3307
QY 2943 T-AAAAAAAAAATGTGAGAAAGTGAATTTTACCAAGGTTTGGCCCAAGGAGGCTT 3001
Db 3308 TAAAAAAAAAATGTGAGAAAGTGAATTTTACCAAGGTTTGGCCCAAGGAGGCTT 3367
```



```

Db 1809 CTTTAACCCCAAGAAAGTGAAGCTGGAAGCGCATATACAGATGCCCTCTCCACAGC 1868
QY 1503 TGCCCGGAGTATGGCAAAAGGTGCGAAGACGTGAACGAACTGCAAACTTAAACAGTGC 1562
Db 1869 TGGCCGGGTATATGGCAAAAGGTGCGAAGACCGTGAACGAACTGCAAACTTAAACAGTGC 1928
QY 1563 AGAAGTCACTGTGCTCTGCTGACCAAAAGCGTCAGATGAAAAATGAGGAAGATGTCAGAAAT 1622
Db 1929 AGAAGTCACTGTGCTCTGCTGACCAAAAGCGTCAGATGAAAAATGAGGAAGATGTCAGAAAT 1988
QY 1623 TATCGGCGACTCTCTTGTAGCCAGCTGCAACAGCGCAAGATTCAGGGAATTTGTACAA 1682
Db 1989 TATCGGCGACTCTCTTGTAGCCAGCTGCAACAGCGCAAGATTCAGGGAATTTGTACAA 2048
QY 1683 GGTGAAGCAGCAGAGCAAGAAATACCTTCAGGAGTCCGCTTCACAGCGCAGAAATGAGG 1742
Db 2049 GGTGAAGCAGCAGAGCAAGAAATACCTTCAGGAGTCCGCTTCACAGCGCAGAAATGAGG 2108
QY 1743 CTCCCAACGAGCAGCAGCAAAACAAAGGATGAAATGAGCCCTTCCAAACCTGACAGAAATG 1802
Db 2109 CTCCCAACGAGCAGCAGCAAAACAAAGGATGAAATGAGCCCTTCCAAACCTGACAGAAATG 2168
QY 1803 AGACCAAAACGACAGCAGCAGATTCGAGCAAAACCAAGACCATCTGAGGAATGAGAAAT 1862
Db 2169 AGACCAAAACGACAGCAGCAGATTCGAGCAAAACCAAGACCATCTGAGGAATGAGAAAT 2228
QY 1863 CTGCGAGAGCGCGCAGAGGACTCTGCGAGAGCCCTGAGAAACCCAGAGGCGGAGAGGCGC 1922
Db 2229 CTGCGAGAGCGCGCAGAGGACTCTGCGAGAGCCCTGAGAAACCCAGAGGCGGAGAGGCGC 2288
QY 1923 GGGGGAAGGTCAACGAGTTTGCAGAAACGAGAGCCGCTCCGCGCCCGCAGGAGCTT 1982
Db 2289 GGGGGAAGGTCAACGAGTTTGCAGAAACGAGAGCCGCTCCGCGCCCGCAGGAGCTT 2348
QY 2289 GGGGGAAGGTCAACGAGTTTGCAGAAACGAGAGCCGCTCCGCGCCCGCAGGAGCTT 2348
Db 1983 CTGCAAGGCTTCAAGCCATCTTCAACATCACTCGGATCTCTCTGAACTCCACAGAGC 2042
QY 2349 CTGCAAGGCTTCAAGCCATCTTCAACATCACTCGGATCTCTCTGAACTCCACAGAGC 2408
Db 2043 CTATCCCTTTTATGTTGAATCAATATGTTGAACGTGTTCAAAGCCCAAGAAATGACAGC 2102
QY 2409 CTATCCCTTTTATGTTGAATCAATATGTTGAACGTGTTCAAAGCCCAAGAAATGACAGC 2468
Db 2103 CCTTTTCTGTGGCAAAATGCTCTGTATACATGTTGTAATATTTAGAAAGGAGAAATGTT 2162
QY 2469 CCTTTTCTGTGGCAAAATGCTCTGTATACATGTTGTAATATTTAGAAAGGAGAAATGTT 2528
Db 2163 AAGATATGTGCGCTGTGTTTACACAGAGGTGCTGACAGCGTAAATATTTTGAATAA 2222
QY 2529 AAGATATGTGCGCTGTGTTTACACAGAGGTGCTGACAGCGTAAATATTTTGAATAA 2588
Db 2223 TATATCAAAATTAATCACTTCACTCAATTTTATCAATTTTATTTTCTTTT 2282
QY 2589 TATATCAAAATTAATCACTTCACTCAATTTTATCAATTTTATTTTCTTTT 2648
Db 2283 AAAGAGAAAGCAAGGCTTTCTAGACTTAAAGATTAAGCTTTGGAGGTCTCAGCGTG 2342
QY 2649 AAAGAGAAAGCAAGGCTTTCTAGACTTAAAGATTAAGCTTTGGAGGTCTCAGCGTG 2708
Db 2343 TAGAGAGAGGCTTTTGAAGCCACCGCAAAATTCACCCAGAGGAAATCTGTCGAGAG 2402
QY 2709 TAGAGAGAGGCTTTTGAAGCCACCGCAAAATTCACCCAGAGGAAATCTGTCGAGAG 2768
Db 2403 GACACTCAGCGAGTCTGTATGATCACTGTGTATGTCAACAGAAAGGATACCGTCTCTTG 2828
QY 2769 GACACTCAGCGAGTCTGTATGATCACTGTGTATGTCAACAGAAAGGATACCGTCTCTTG 2828
Db 2463 AAAGAGAAATCTGTCACTCTCATGCTGCTAGCTCATACACCCATTTCTCTTGCTT 2522
QY 2829 AAAGAGAAATCTGTCACTCTCATGCTGCTGCTAGCTCATACACCCATTTCTCTTGCTT 2888
Db 2523 CACAGGTTTAACTGGTTTGTGCAATATGCTATATATATTTCTGCTCTCTGTTTAA 2582
QY 2889 CACAGGTTTAACTGGTTTGTGCAATATGCTATATATATTTCTGCTCTCTGTTTAA 2948

```

```

QY 2583 TCTCTCCCTCCCTCCCTCCCTCTCTCTCTCATCTCCATCTTTTGAATTTCTCATCC 2642
Db 2949 TCTCTCCCTCCCTCCCTCCCTCCCTCTCTCTCTCATCTCCATCTTTTGAATTTCTCATCC 3008
QY 2643 CTCATCTCAATCCCGTATCTTACGACACCCCGCCCCCCCCCAGGAAAGAGTCTGTAGT 2702
Db 3009 CTCATCTCAATCCCGTATCTTACGCA--CCCCCCCCCAGGAAAGAGTCTGTAGT 3067
QY 2703 ATCAATCAACAAAGGAAACAAAGGAAACACACAAACAGCCTCACTTACACTTGG 2762
Db 3068 ATCAATCAACAAAGGAAACAAAGGAAACACACAAACAGCCTCACTTACACTTGG 3127
QY 2763 TTAACAAAGAAACAGATCAATGTATCTTGTCTTACGCTTTTGAAGAGAAACAGG 2822
Db 3128 TTAACAAAGAAACAGATCAATGTATCTTGTCTTACGCTTTTGAAGAGAAACAGG 3187
QY 2823 AACCAACCAACCAACCAATCAATCAACCAACCAAGAAATTTCCAAATGAAGAAATGAT 2882
Db 3188 AACCAACCAACCAACCAATCAATCAACCAACCAAGAAATTTCCAAATGAAGAAATGAT 3247
QY 2883 TTTGCTTTTGAATTTTGTGTATPACCATCATATTTACAGCAAAATGATCTTCTT 2942
Db 3248 TTTGCTTTTGAATTTTGTGTATPACCATCATATTTACAGCAAAATGATCTTCTT 3307
QY 2943 T-AAAAAAAAAATGTGAGAGAAAGTAAATTTTACAAAGTTTGGCCCAAGGCGTTA 3001
Db 3308 TAAAAAATAAATGTGAGAGAAAGTAAATTTTACAAAGTTTGGCCCAAGGCGTTA 3367
QY 3002 AATTCAAGATTTTAAACAGAAACCAACAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3061
Db 3368 AATTCAAGATTTTAAACAGAAACCAACAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3427
QY 3062 CTCAGACCTTGTCTGTGTGTTCCCTTACAGATTTTGAAGTAAAGTAAAGTAAAGTAAAG 3121
Db 3428 CTCAGACCTTGTGTGTGTTCCCTTACAGATTTTGAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3487
QY 3122 TTTTATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3181
Db 3488 TTTTATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3547
QY 3182 AAGAGTGAACAGTCCAGTGTGCAACAGTGTGCAATGTTCTCCGCTAGCAAGAG 3241
Db 3548 AAGAGTGAACAGTCCAGTGTGCAACAGTGTGCAATGTTCTCCGCTAGCAAGAG 3507
QY 3242 CNAATAGGCTCTTTTGAACAAACCTTGAATGTTTATTT 3283
Db 3608 CTATATGGCTCTTTTGAACAAACCTTGAATGTTTATTT 3649

```

```

RESULT 5
US-09-899-651-6
/ Sequence 6, Application US/09899651
/ Patent No. US2002011470A1
/ GENERAL INFORMATION:
/ APPLICANT: Chen, Yao-Tseng
/ APPLICANT: Gure, Ali
/ APPLICANT: Tseng, Solam
/ APPLICANT: Stockert, Elisabeth
/ APPLICANT: Jager, Mike
/ APPLICANT: Knuth, Alexander
/ APPLICANT: Old, Lloyd J.
/ TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
/ TITLE OF INVENTION: Associated
/ TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
/ FILE REFERENCE: LUD 5538
/ CURRENT APPLICATION NUMBER: US/09/899,651
/ PRIOR FILING DATE: 2001-07-06
/ PRIOR APPLICATION NUMBER: US/09/061,709
/ NUMBER OF SEQ ID NOS: 8
/ SEQ ID NO 6
/ LENGTH: 3412

```



```
Db 2041 CGAGGAGGCGGGAGAGCTCAGCCAGTTTCCGAAACCAACGACCCCGCTCCGCC 2100
Qy 1972 CCCAGGAGCTTCTGAGGCTTCAAGCCATCCACTTCAACATCCATCGGATCTCTCGAA 2031
Db 2101 CCCAGGAGCTTCTGAGGCTTCAAGCCATCCACTTCAACATCCATCGGATCTCTCGAA 2160
Qy 2032 CTCCACAGAGCTATCCCTTTTAACTTGAATCAATAGTGAAGCTGTTCAAGGCAAGC 2091
Db 2161 CTCCACAGAGCTATCCCTTTTAACTTGAATCAATAGTGAAGCTGTTCAAGGCAAGC 2220
Qy 2092 AAAATGACACCCCTTTTCTGTGCAAAACGTCTCTGTAACATGTGTGTAATATTGAAA 2151
Db 2221 AAAATGACACCCCTTTTCTGTGCAAAACGTCTCTGTAACATGTGTGTAATATTGAAA 2280
Qy 2152 GGGAGATGTTAAGATATGTGAGCTGTGTGTTACACAGGAGCTGTGACGGGTATATAT 2211
Db 2281 GGGAGATGTTAAGATATGTGAGCTGTGTGTTACACAGGAGCTGTGACGGGTATATAT 2340
Qy 2212 TTTAAGAAATATATATCAATTAATCACTCACTCACTCACTCACTCACTCACTCACTCA 2271
Db 2341 TTTAAGAAATATATATCAATTAATCACTCACTCACTCACTCACTCACTCACTCACTCA 2400
Qy 2272 TTTTCTTTTAAAGAAAGAGGCTTTTCTAGAATTAAAGAAATAAAGTCTTTGGAG 2331
Db 2401 TTTTCTTTTAAAGAAAGAGGCTTTTCTAGAATTAAAGAAATAAAGTCTTTGGAG 2460
Qy 2332 GTCTACGCTGTAGAGAGAGCTTTGAGGCAACCGGCAAAATTTACCCAGAGGAAT 2391
Db 2461 GTCTACGCTGTAGAGAGAGCTTTGAGGCAACCGGCAAAATTTACCCAGAGGAAT 2520
Qy 2392 CTCGTCGGAAGAGACCTCAGGAGCTGTGATGATCACTGTATGTCAACAGAGGATA 2451
Db 2521 CTCGTCGGAAGAGACCTCAGGAGCTGTGATGATCACTGTATGTCAACAGAGGATA 2580
Qy 2452 CCGTCTCTTGAAGAGAACTCTGTCACTCTCTCACTCTCTCTCTCTCTCTCTCTCTCT 2511
Db 2581 CCGTCTCTTGAAGAGAACTCTGTCACTCTCTCACTCTCTCTCTCTCTCTCTCTCTCT 2640
Qy 2512 TCTCTTGTCTTCAAGAGTTTAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2571
Db 2641 TCTCTTGTCTTCAAGAGTTTAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2700
Qy 2572 CTCTCTGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2631
Db 2701 CTCTCTGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760
Qy 2632 TTTCTCATCTCTCTCACTCAATCCGTATCTAGCAGCCGCCGCCGCCGCCGCCGCCGCC 2691
Db 2761 TTTCTCATCTCTCTCACTCAATCCGTATCTAGCAGCCGCCGCCGCCGCCGCCGCCGCC 2820
Qy 2692 GTGCTGTAGATCAATCAACACAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAA 2751
Db 2821 GTGCTGTAGATCAATCAACACAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAA 2880
Qy 2752 CTTAACCTTGATCTCAAAAGAAAGAGTCAATGGTACTTGTCTAGAGTTTGGAAAG 2811
Db 2881 CTTAACCTTGATCTCAAAAGAAAGAGTCAATGGTACTTGTCTAGAGTTTGGAAAG 2940
Qy 2812 AGGAAACAGGAAACCAACCAACCAATCAACCAACCAACCAACCAACCAACCAACCAATG 2871
Db 2941 AGGAAACAGGAAACCAACCAACCAATCAACCAACCAACCAACCAACCAACCAACCAATG 3000
Qy 2872 AAGAAATGATTTTGTCTTTTGTGATTTGTGTATAGCCATCAATTCACCAAAATG 2931
Db 3001 AAGAAATGATTTTGTCTTTTGTGATTTGTGTATAGCCATCAATTCACCAAAATG 3060
Qy 2932 ATTCCCTTCTTTAAAAAATATGTGAGAAAGTGAATTTTACCAAGTGTGTGCC 2991
Db 3061 ATTCCCTTCTTTAAAAAATATGTGAGAAAGTGAATTTTACCAAGTGTGTGCC 3120
Qy 2992 CAGGAGGATTAATTCACAGATTTTAAAGAGAAACACAGAGAAAGTACTCAG 3051
```

```
Db 3121 CAGGAGGATTAATTCACAGATTTTAAAGAGAAACACAGAGAAAGTACTCAG 3180
Qy 3052 GTGTTTAACTCTAGACACTTGTCTGTGTCTTCTCTTAAAGATTTTGAAGCTGATAG 3111
Db 3181 GTGTTTAACTCTAGACACTTGTCTGTGTCTTCTCTTAAAGATTTTGAAGCTGATAG 3240
Qy 3112 TTGAGCATTTTAAATTTTAAATTAATAAATAGTTGAAATTAAGATATCACT 3171
Db 3241 TTGAGCATTTTAAATTTTAAATTAATAAATAGTTGAAATTAAGATATCACT 3300
Qy 3172 GCCAGCTGTGAGAGAGTCAAGTCCAGATGTGCAACAGCTGTGCAATGTCTCCGCT 3231
Db 3301 GCCAGCTGTGAGAGAGTCAAGTCCAGATGTGCAACAGCTGTGCAATGTCTCCGCT 3360
Qy 3232 AGCCAAAGACNAATATGAGCTCTTTTGAACAAACCTTGAATTTATTT 3283
Db 3361 AGCCAAAGACNAATATGAGCTCTTTTGAACAAACCTTGAATTTATTT 3412

RESULT 6
US-09-764-864-329
; Sequence 329, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT FILING DATE: US/09/764,864
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: Patencin Ver. 2.0
; SEQ ID NO 329
; LENGTH: 3694
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-329

Query Match 93.0%; Score 3052.6; DB 9; Length 3694;
Best Local Similarity 96.0%; Pred. No. 0;
Matches 3214; Conservative 1; Mismatches 1; Indels 132; Gaps 3;
```

QY 548 TAGCTCCCTTTCGCCCCCTCAGCAGAGCCCGAGGCCACTTTCGCCGGAGCAAG 607
Db 571 TAGCTCCCTTTCGCCCCCTCAGCAGAGCCCGAGGCCACTTTCGCCGGAGCAAG 630
QY 608 GCCAGGCCCTTGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGGATCTGG 667
Db 631 GCCAGGCCCTTGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGGATCTGG 690
QY 668 TCCCCACCCAGTTTGTGGTGCCATCATCTGGAAGAAGGGGCTTTCACATTAAGAACATCA 727
Db 691 TCCCCACCCAGTTTGTGGTGCCATCATCTGGAAGAAGGGGCTTTCACATTAAGAACATCA 750
QY 728 CTAAACAGACCCAGTCCCGGGTAGATATCATAGAAAAGAACTCTGAGAGCTGCAGAGA 787
Db 751 CTAAACAGACCCAGTCCCGGGTAGATATCATAGAAAAGAACTCTGAGAGCTGCAGAGA 810
QY 788 AGCTGTACCAATCCATGCCACCCCGAGGGGACTTTCGAAGCATGCCGATGATTTCTTG 847
Db 811 AGCTGTACCAATCCATGCCACCCCGAGGGGACTTTCGAAGCATGCCGATGATTTCTTG 870
QY 848 AATCATGACAGAAAGAGGAGATAGAGCAAACTTTCGAGAGATTTCTCTGAAATCT 907
Db 871 AATCATGACAGAAAGAGGAGATAGAGCAAACTTTCGAGAGATTTCTCTGAAATCT 930
QY 908 TGGCACAATGAGCTTGGTGGAAAGCTGATTGGAAAAGAGGAGCAAAATTGAAAGAAA 967
Db 931 TGGCACAATGAGCTTGGTGGAAAGCTGATTGGAAAAGAGGAGCAAAATTGAAAGAAA 990
QY 968 TTGAACATGAAACAGGAGCCAAAGTAAATCTCATCTTTGACAGATTGAGCATATACA 1027
Db 991 TTGAACATGAAACAGGAGCCAAAGTAAATCTCATCTTTGACAGATTGAGCATATACA 1050
QY 1028 ACCGGAAGAGACATCATCTGTGAAGGGCAAGTTGAGGCTGTGCCAGTCTGATAG 1087
Db 1051 ACCGGAAGAGACATCATCTGTGAAGGGCAAGTTGAGGCTGTGCCAGTCTGATAG 1110
QY 1088 AGATTATGAAGAGAGCTGCTGAGAGGCTTTGAAATGATATGCTGGCTGTTAAC----- 1140
Db 1111 AGATTATGAAGAGAGCTGCTGAGAGGCTTTGAAATGATATGCTGGCTGTTAACCAACAG 1170
QY 1141 ----- 1140
Db 1171 CCAATCTGATCCAGGGTGAACCTCAGCGCACTTGGCATCTTTCAACAGAGACTGTCCG 1230
QY 1141 ----- 1140
Db 1231 TGCATATCCACAGCAGGGGCCCGGAGCTCCCCCGCTGCCCTACACCCCTTCA 1290
QY 1141 --ACCACTCCGATACTTCTCCAGCTGTACCCCATCAAGTTTGGCCGTTCCGCG 1198
Db 1291 CTACCCACTCCGATACTTCTCCAGCTGTACCCCATCAAGTTTGGCCGTTCCGCG 1350
QY 1199 ATCATCACTTTTCCAGAGCAGAGATGTGAATCTTTCAATCCAAACCAAGCTGTGG 1258
Db 1351 ATCATCACTTTTCCAGAGCAGAGATGTGAATCTTTCAATCCAAACCAAGCTGTGG 1410
QY 1259 GCGCATCATCGGGAAGAGAGGGGCAACATCAAAAGCTGGGAGAGTTGGCGAGGCT 1318
Db 1411 GCGCATCATCGGGAAGAGAGGGGCAACATCAAAAGCTGGGAGAGTTGGCGAGGCT 1470
QY 1319 CTATCAAGATTGGCCCTGCGGAAGGCCAGACGTCAGCGAAAGATGGTCAATCAACCG 1378
Db 1471 CTATCAAGATTGGCCCTGCGGAAGGCCAGACGTCAGCGAAAGATGGTCAATCAACCG 1530
QY 1379 GCGCACCGGAAGCCAGTTCAAGGCCAGAGGAGATCTTTGGGAAACTGAAAGAGAAA 1438
Db 1531 GCGCACCGGAAGCCAGTTCAAGGCCAGAGGAGATCTTTGGGAAACTGAAAGAGAAA 1590
QY 1439 ACTTCTTTAAACCCCAAGAGAGTGAAGTGAAGCCATATCAAGATGCTCTTCCA 1498
Db 1591 ACTTCTTTAAACCCCAAGAGAGTGAAGTGAAGCCATATCAAGATGCTCTTCCA 1650

QY 1499 CAGCTGGCCGGGTGATTGGCAAGGTGGCAAGACCGTGAACGAACTGCAAGAACTTAACCA 1558
Db 1651 CAGCTGGCCGGGTGATTGGCAAGGTGGCAAGACCGTGAACGAACTGCAAGAACTTAACCA 1710
QY 1559 GTGCAGAAAGTCACTGTGCTGTGACCAAAAGCCAGATGAAATAGGAAGTATGCTCA 1618
Db 1711 GTGCAGAAAGTCACTGTGCTGTGACCAAAAGCCAGATGAAATAGGAAGTATGCTCA 1770
QY 1619 GAATTATGGGCACTTCTTTGCTAGCCAGACTGCAACAGCGCAAGATCAGGAAATTTGAC 1678
Db 1771 GAATTATGGGCACTTCTTTGCTAGCCAGACTGCAACAGCGCAAGATCAGGAAATTTGAC 1830
QY 1679 AACAGTGAACAGACAGAGAGCAAGAAATTCCTCAGGAGGTGCTCTCAAGCGCAGCAAGT 1738
Db 1831 AACAGTGAACAGACAGAGAGCAAGAAATTCCTCAGGAGGTGCTCTCAAGCGCAGCAAGT 1890
QY 1739 GAGGCTCCCAAGGACCAAGCAAAACAAAGATGATAGCCCTTCCAAACCTGACAG 1798
Db 1891 GAGGCTCCCAAGGACCAAGCAAAACAAAGATGATAGCCCTTCCAAACCTGACAG 1950
QY 1799 AATGAGACCAACGACGACGACGATGGGAGCAAAACCAAGACATCTGAGAGATGAG 1858
Db 1951 AATGAGACCAACGACGACGACGATGGGAGCAAAACCAAGACATCTGAGAGATGAG 2010
QY 1859 AAGTCTGCGAGAGGCGGCAAGGACTTTCGCGAGGCCCTGAGAACCCAGGGGCGAGAG 1918
Db 2011 AAGTCTGCGAGAGGCGGCAAGGACTTTCGCGAGGCCCTGAGAACCCAGGGGCGAGAG 2070
QY 1919 GGGCGGGGAAAGGTGAGCCAGGTTTGGCAAGAACCAAGAGCCCGCTCCGCGCCCGAG 1978
Db 2071 GGGCGGGGAAAGGTGAGCCAGGTTTGGCAAGAACCAAGAGCCCGCTCCGCGCCCGAG 2130
QY 1979 GCTTCTGAGGCTTCAAGCATCACTTCAACCATCACTCGATCTCTGAACTCCAC 2038
Db 2131 GCTTCTGAGGCTTCAAGCATCACTTCAACCATCACTCGATCTCTGAACTCCAC 2190
QY 2039 GAGCTATCCCTTTAATTGAACATATAGTGAAGTGTTCAAAGCCAAAGAAATGC 2098
Db 2191 GAGCTATCCCTTTAATTGAACATATAGTGAAGTGTTCAAAGCCAAAGAAATGC 2250
QY 2099 ACAACCTTTTCTGTGCAAAATGCTCTGTAGATGATGTATATTAAGAAAGGAA 2158
Db 2251 ACAACCTTTTCTGTGCAAAATGCTCTGTAGATGATGTATATTAAGAAAGGAA 2310
QY 2159 TGTTAAGATATGAGCCTGTGGTTACAGAGGTGCTGCAAGCGTATATATTTAGAA 2218
Db 2311 TGTTAAGATATGAGCCTGTGGTTACAGAGGTGCTGCAAGCGTATATATTTAGAA 2370
QY 2219 ATATATATCAAAATCACTCACTTCAATTTTATCAATTAATTTTCT 2278
Db 2371 ATATATATCAAAATCACTCACTTCAATTTTATCAATTAATTTTCT 2430
QY 2279 TTTTAAAGAGAGAGGCTTTCTAGACTTTTAAAGATTAAGCTTGGAGGTCTCAC 2338
Db 2431 TTTTAAAGAGAGAGGCTTTCTAGACTTTTAAAGATTAAGCTTGGAGGTCTCAC 2490
QY 2339 GGTGTAGAGAGGCTTTGAGGCCACCCGCAAAATTCACCAAGAGAAATCTGTCG 2398
Db 2491 GGTGTAGAGAGGCTTTGAGGCCACCCGCAAAATTCACCAAGAGAAATCTGTCG 2550
QY 2399 GAAGGACATCAAGGAGATTTGATCACTGTGTATGTCAAGAAAGGATACCGTCTC 2458
Db 2551 GAAGGACATCAAGGAGATTTGATCACTGTGTATGTCAAGAAAGGATACCGTCTC 2610
QY 2459 CTTGAAGAGAACTCTGTACATCCCATGCGGTCTAGCTCAACCACTTCTCTT 2518
Db 2611 CTTGAAGAGAACTCTGTACATCCCATGCGGTCTAGCTCAACCACTTCTCTT 2670
QY 2519 GCTTCAAGGATTTTAAACTGTGTTTGGCATAGCTATATTAATCTCTGCTCTCTG 2578
Db 2671 GCTTCAAGGATTTTAAACTGTGTTTGGCATAGCTATATTAATCTCTGCTCTCTG 2730
QY 2579 TTTATCTCTCCCTCCCTCCCTCTTCTCTCATCTCTTGTGAATTTCTC 2638

Db	2731	TTTATCTCTCCCTCCCTCCCTCCCTCTTCTTCATCTCAATCTTTTGAAATTTCTTC	2790
QY	2639	ATCCCTCACTCAATCCGATCTCTAGCA--CCCCCCCCCCCCGAGCAAAAGCACTGCT	2696
Db	2791	ATCCCTCACTCAATCCGATCTAGCAACCCCCCCCCCCCCGAGCAAAAGCACTGCT	2850
QY	2697	CTGAGTATCAATCAACAAAAGGAACAAAAGGAACACACAAACAGCTTCACTTAC	2756
Db	2851	CTGAGTATCAATCAACAAAAGGAACAAAAGGAACACACAAACAGCTTCACTTAC	2910
QY	2757	ACTTGATTCTCAAAAAGAAAGAGTCAATGTATCTTGTCTTACGCTTTTGGAAAGGAA	2816
Db	2911	ACTTGATTCTCAAAAAGAAAGAGTCAATGTATCTTGTCTTACGCTTTTGGAAAGGAA	2970
QY	2817	AACAGAAACCCCAACCAACCAATCAACCAACCAAAAGAAAATTTCCCAATGAAGA	2876
Db	2971	AACAGAAACCCCAACCAACCAATCAACCAACCAAAAGAAAATTTCCCAATGAAGA	3030
QY	2877	ATGTAATTTGTCTTTTGGCAATTTTGGTATTAAGCCATCAATTTGACAAATGTATTC	2936
Db	3031	ATGTAATTTGTCTTTTGGCAATTTTGGTATTAAGCCATCAATTTGACAAATGTATTC	3090
QY	2937	TTTCTTTT-AAAAAATAATGTGAGGAAGATAGAAATTTACCAAGTTGTTGGCCACG	2995
Db	3091	TTTCTTTTAAAAAATAATGTGAGGAAGATAGAAATTTACCAAGTTGTTGGCCACG	3150
QY	2996	GCGTAAATTTACAGATTTTTTTTAAACGAAACAAACAGAGAACCTACCTAGGTGT	3055
Db	3151	GCGTAAATTTACAGATTTTTTTTAAACGAAACAAACAGAGAACCTACCTAGGTGT	3210
QY	3056	TTTAAACCTCAGACCTTGCTGTGTGTTCCCTTAGAGATTTTGAAGCTGATGATGTG	3115
Db	3211	TTTAAACCTCAGACCTTGCTGTGTGTTCCCTTAGAGATTTTGAAGCTGATGATGTG	3270
QY	3116	AGCATTTTTTTATTTTTTTAATAAAATGAGTGGAAAAAATAAAGATACACCTGCCA	3175
Db	3271	AGCATTTTTTTATTTTTTTAATAAAATGAGTGGAAAAAATAAAGATACACCTGCCA	3330
QY	3176	GCGTGAAGAGGTGACAGTCCAGAGTGTGCAACAGCTGTTTGAATTTGTTCCGCTAGCC	3235
Db	3331	GCGTGAAGAGGTGACAGTCCAGAGTGTGCAACAGCTGTTTGAATTTGTTCCGCTAGCC	3390
QY	3236	AAGAACCAATATGAGCTTCTTTGGACAAACCTTGAAAAATGTTTATTTT 3283	
Db	3391	AAGAACCAATATGAGCTTCTTTGGACAAACCTTGAAAAATGTTTATTTT 3438	
RESULT 7			
US-09-764-864-332			
; Sequence 332, Application US/09764864			
; Patent No. US20020132753A1			
; GENERAL INFORMATION:			
; APPLICANT: Rosen et al.			
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies			
; FILE REFERENCE: PT223			
; CURRENT APPLICATION NUMBER: US/09/764,864			
; PRIORITY FILING DATE: 2001-01-17			
; NUMBER OF SEQ ID NOS: 1792			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 332			
; LENGTH: 2290			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-09-764-864-332			

Query Match	60.6%;	Score 1969.8;	DB 9;	Length 2290;
Best Local Similarity	99.4%;	Pred. No. 0;		
Matches 2078;	Conservative	0;	Mismatches 8;	Indels 4;
				Gaps 3.
QY	1247	CCGAGCTGTGGGCGCCATCATCGGGAAGAAAGGGGACACATCAACAGCTGGCGAGAT		1300

Db	25	CCCAAGCCTCCGCGCCCATTCATATGGGAAATAAGGGGCGACACATTCAAACAGCTGGCGAGAT	84
QY	1307	TCGCGGAGAGCTCTATCAAGATTGCCCTTCGCGAAGGCCAGACGTCAGCGAAAGATGG	1366
Db	85	TCGCGGAGAGCTCTATCAAGATTGCCCTTCGCGAAGGCCAGACGTCAGCGAAAGATGG	144
QY	1367	TCATCATCACCGGCGCACCGGAAAGCCCATGTTCAAGGCCAGAGGACGGATCTTTGGGAAAC	1426
Db	145	TCATCATCACCGGCGCACCGGAAAGCCCATGTTCAAGGCCAGAGGACGGATCTTTGGGAAAC	204
QY	1427	TGAAGAAGAAAACCTTCTTAAACCCAAAGAAAGATGAAAGCTGAAAGCGCATATCAGAG	1486
Db	205	TGAAGAAGAAAACCTTCTTAAACCCAAAGAAAGATGAAAGCTGAAAGCGCATATCAGAG	264
QY	1487	TGCCCTCTTCACAGCTGCGCGGATGATTGCGAAAGGTGCAAGACCCGTGAACGAACTCG	1546
Db	265	TGCCCTCTTCACAGCTGCGCGGATGATTGCGAAAGGTGCAAGACCCGTGAACGAACTCG	324
QY	1547	AGAACTTACCGATGCGAAGATCATCGTGCCTGTGACCAAAACGGCAGATGAAATAGAG	1606
Db	325	AGAACTTACCGATGCGAAGATCATCGTGCCTGTGACCAAAACGGCAGATGAAATAGAG	384
QY	1607	AAGATGATGTGAGAAATATCGGGACCTCTTTCTAGCCAGACATGCGACGCGAAGATCA	1666
Db	385	AAGATGATGTGAGAAATATCGGGACCTCTTTCTAGCCAGACATGCGACGCGAAGATCA	444
QY	1667	GGAAATTTGTACAAACAGGTGAAGCGACGAGACGAAATTCCTTCAGGAGTGGCTTCAC	1726
Db	445	GGAAATTTGTACAAACAGGTGAAGCGACGAGACGAAATTCCTTCAGGAGTGGCTTCAC	504
QY	1727	AGCGGACAAATGAGGCTCCACAGGACCAAGCAAAACAAACGAGTGAATGTAGCCCTTC	1786
Db	505	AGCGGACAAATGAGGCTCCACAGGACCAAGCAAAACAAACGAGTGAATGTAGCCCTTC	564
QY	1787	AACACCTGACAGATGAGACCAAAACGACGACGACGATCGGAGCAAAACAAAGACAT	1846
Db	565	AACACCTGACAGATGAGACCAAAACGACGACGACGATCGGAGCAAAACAAAGACAT	624
QY	1847	CTGAGAAATGAGAAAGTCTGCGGAGCGCGCCAGGACCTCTGCGAGGCTCTTGAAACCCCA	1906
Db	625	CTGAGAAATGAGAAAGTCTGCGGAGCGCGCCAGGACCTCTGCGAGGCTCTTGAAACCCCA	684
QY	1907	GGGGGCCGAGAGGGCGGGGAAAGGTCAGACAGTTTCCAGAACCAACCGAGCCCGCTC	1966
Db	685	GGGGGCCGAGAGGGCGGGGAAAGGTCAGACAGTTTCCAGAACCAACCGAGCCCGCTC	744
QY	1967	CGGCCCCCGAGGCTTCTGACAGGCTTCAGCCATCCATTCACATTCGATCTCTC	2026
Db	745	CGGCCCCCGAGGCTTCTGACAGGCTTCAGCCATCCATTCGATCTCTC	804
QY	2027	CTGAACCTCCACGACGCTATCCCTTTTAGTTGAACATAATAGTGAACGTTTCAAGC	2086
Db	805	CTGAACCTCCACGACGCTATCCCTTTTAGTTGAACATAATAGTGAACGTTTCAAGC	864
QY	2087	CAAGCAAAATGACACCCCTTTTCGTGAGCAAAATGCTCTGTAACATGTGTACATAT	2146
Db	865	CAAGCAAAATGACACCCCTTTTCGTGAGCAAAATGCTCTGTAACATGTGTACATAT	924
QY	2147	AGAAAGGAAAGATGTTAAGATATGTGCGCTGTGGTTTCAACAGGATGCTGACGCGTAA	2206
Db	925	AGAAAGGAAAGATGTTAAGATATGTGCGCTGTGGTTTCAACAGGATGCTGACGCGTAA	984
QY	2207	TATATTTAGAAATAATATATCAATAAATCACTCAATTTTAACTATTTAATCAATTTTA	2266
Db	985	TATATTTAGAAATAATATATCAATAAATCACTCAATTTTAACTATTTAATCAATTTTA	1044
QY	2267	ATTTTCTTTCTTTTAAAGAAAGCGCTTCTAGCTTTTAAAGATTTAAAGTCTTT	2326
Db	1045	ATTTTCTTTCTTTTAAAGAAAGCGCTTCTAGCTTTTAAAGATTTAAAGTCTTT	1104
QY	2327	GGGAGGCTCACGGGTGAGAGAGCTTTGAGGCCACCCGACAAAATTCAACCAAGG	2386
Db	1105	GGGAGGCTCACGGGTGAGAGAGCTTTGAGGCCACCCGACAAAATTCAACCAAGG	1164

```

OY 2387 GAAATCTGTCGGAAGGACACTCAGGCGAGTCTGGATCACTGTGATGTCACAGAG 2446
DB 1165 GAAATCTGTCGGAAGGACACTCAGGCGAGTCTGGATCACTGTGATGTCACAGAG 1224
OY 2447 GATACCGTCTCTTGAAGAGAACTCTGTCACTCTCTGATGCTGTAGCTCATAC 2506
DB 1225 GATACCGTCTCTTGAAGAGAACTCTGTCACTCTCTGATGCTGTAGCTCATAC 1284
OY 2507 CCATTTCTTTGCTTCAAGGTTTAACTGGTTTTTGAATAGTATGATATATTTCTC 2566
DB 1285 CCATTTCTTTGCTTCAAGGTTTAACTGGTTTTTGAATAGTATGATATATTTCTC 1344
OY 2567 TGTCTCTCTGTTTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2626
DB 1345 TGTCTCTCTG- TTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1403
OY 2627 TTGAATTTCTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2684
DB 1404 TTGAATTTCTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1463
OY 2685 CAAAGCAGTGTCTGAGATATCATCATCAAAAAGAAAGAAAGGAAACACACAAACA 2744
DB 1464 CAAAGCAGTGTCTGAGATATCATCATCAAAAAGAAAGAAAGGAAACACACAAACA 1523
OY 2745 GCCTCAACTTACACTTGTGTTACTCAAAAAGAAAGAAAGTCAATGTTCTCTAGGTT 2804
DB 1524 GCCTCAACTTACACTTGTGTTACTCAAAAAGAAAGAAAGTCAATGTTCTCTAGGTT 1583
OY 2805 TTGGAAGAGAAACAGAACCCACCAACCAACCAACCAACCAACCAACCAACCAAC 2864
DB 1584 TTGGAAGAGAAACAGAACCCACCAACCAACCAACCAACCAACCAACCAACCAAC 1643
OY 2865 CACAATGAAAGATGATTTTGTCTTTTGAATTTTGTGATTAAGCATTAATTCAG 2924
DB 1644 CACAATGAAAGATGATTTTGTCTTTTGAATTTTGTGATTAAGCATTAATTCAG 1703
OY 2925 CAAAATGATTCCTTCTTTT-AAAAAATAATGTGAGAGAAAGTAAATTTACAGT 2983
DB 1704 CAAAATGATTCCTTCTTTTAAAAAATAATGTGAGAGAAAGTAAATTTACAGT 1763
OY 2984 TGTGGCCCAAGGCGTTAAATTCACAGATTTTAAAGAGAAACACACAGAGAGC 3043
DB 1764 TGTGGCCCAAGGCGTTAAATTCACAGATTTTAAAGAGAAACACACAGAGAGC 1823
OY 3044 TACCTCAGGTGTTTAACTCAGACCTGCTGTGTTTCTCTTGAATTTTGA 3103
DB 1824 TACCTCAGGTGTTTAACTCAGACCTGCTGTGTTTCTCTTGAATTTTGA 1883
OY 3104 GCTGATAGTTGAGCATTTTATTTTATTTTAAATGAATGAGTGAATAATAAGA 3163
DB 1884 GCTGATAGTTGAGCATTTTATTTTATTTTAAATAATAATAATAATAATAAGA 1943
OY 3164 TATCAACTGCGCAGCTGAGAGAGTGAACGTCAAGTGTCAACAGCTGTTCTGAATGT 3223
DB 1944 TATCAACTGCGCAGCTGAGAGAGTGAACGTCAAGTGTCAACAGCTGTTCTGAATGT 2003
OY 3224 CTTCCTGTCAGCAAGAACATATAGCTCTTTTGAACAACTTTGAAATGTTATTT 3283
DB 2004 CTTCCTGTCAGCAAGAACATATATAGCTCTTTTGAACAACTTTGAAATGTTATTT 2063

```

RESULT 8

```

US-10-262-445-39
; Sequence 39, Application US/10262445
; Publication No. US20040014058A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook II, John
; APPLICANT: Burgess, Catherine
; APPLICANT: Gatterton, Elina
; APPLICANT: Chant, John
; APPLICANT: Chaudhuri, Amitabha
; APPLICANT: Edinger, Shlomit

```

```

; APPLICANT: Gerlach, Valerie
; APPLICANT: Girot, Loic
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Kexuda, Ramesh
; APPLICANT: Mezes, Peter
; APPLICANT: Millet, Isabelle
; APPLICANT: Ooi, Chean Eng
; APPLICANT: Paturajan, Meera
; APPLICANT: Rieger, Daniel
; APPLICANT: Spytek, Kimberly
; APPLICANT: Taupier Jr., Raymond J.
; APPLICANT: Zehrusen, Bryan
; APPLICANT: Zhong, Haihong
; APPLICANT: Zhong, Wei
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYPEPTIDES ENCODING THEM AND METHODS
; FILE REFERENCE: 21402-462D
; CURRENT APPLICATION NUMBER: US/10/262,445
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/327,454
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 60/327,917
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,849
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/329,414
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 60/330,142
; PRIOR FILING DATE: 2001-10-17
; PRIOR APPLICATION NUMBER: 60/341,058
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: 60/343,629
; PRIOR FILING DATE: 2001-10-24
; PRIOR APPLICATION NUMBER: 60/349,575
; PRIOR FILING DATE: 2001-10-29
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 133
; SOFTWARE: CuroSeqdist version 0.1
; SEQ ID NO 39
; LENGTH: 1707
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5)..(1669)
US-10-262-445-39

Query Match 47.6%; Score 1561.2; DB 16; Length 1707;
Best Local Similarity 95.0%; Pred. No. 0;
Matches 1624; Conservative 0; Mismatches 83; Indels 3; Gaps 1;

OY 66 ACGATGATGACAAAGCTTTTACATCGGGAACCTGAGCCCGCCGTCAACCGCCGAGACT 125
DB 1 ACGATGATGACAAAGCTTTTACATCGGGAACCTGAGCCCGCCGTCAACCGCCGAGACT 60

OY 126 CCGGAGCTCTTTGGGAGACAGAGCTGCCCCCTGGGGGACAGGTCTGCTGAAGTCCG 185
DB 61 CCGGAGCTCTTTGGGAGACAGAGCTGCCCCCTGGGGGACAGGTCTGCTGAAGTCCG 120

OY 186 CTACGCTTGTGAGACTACCCGACAGAACTGGCCATCGGCGCATCGAGACCCCTCTC 245
DB 121 CTACGCTTGTGAGACTACCCGACAGAACTGGCCATCGGCGCATCGAGACCCCTCTC 180

OY 246 GGGTAAAGTGAATTCATGAGGAAATTCATGAAAGTTGATTAATCTCTTAAAAAGCT 305
DB 181 GGGTCAAGTGAATTCATGAGGAAATTCATGAAAGTTGATTAATCTCTTAAAAAGCT 240

OY 306 AAGGACGAGAAATTCAGATTGAAACATCCCTCTCTACCTGCAATGGAGGTGTTGA 365

```

Db 241 AAGAGGAGGAACTTCGATTCGAAATATCCCTCTCACTGCAAGTGGAGTGTGGA 300
QY 366 TGAATTTTGGCTCAATATGGAACAGTGGAGATGTGAAACAGTCAACACAGACAGA 425
Db 301 TGAATTTTGGCTCAATATGGAACAGTGGAGATGTGAAACAGTCAACACAGACAGA 360
QY 426 AACCCTGTTGCAACCTGCACTATGCAACAGAAAGAAAGAAATATGAGCATGAGAA 485
Db 361 GACCCCTGTTGCAACCTGCACTATGCAACAGAAAGAAAGAAATATGAGCATGAGAA 420
QY 486 GCTAAGCGGAGCATGATTTGAGAACTAATCTCTTCAAGATTTCTCAATCCCGATGAGA 545
Db 421 GCTAAGCGGAGCATGATTTGAGAACTAATCTCTTCAAGATTTCTCAATCCCGATGAGA 480
QY 546 GGTAGCTCCCTTGGCCCTTCAAGAGCCAGCGTGGGAGCACTCTTCCCGGAGCA 605
Db 481 GGTAGCTCCCTTGGCCCTTCAAGAGCCAGCGTGGGAGCACTCTTCCCGGAGCA 540
QY 606 AGGCCAGCCCTTGGGAGGCACTTCAAGGCGCAGAGATGATTTCCCGCTGCGATCCT 665
Db 541 AGGCCAGCCCTTGGGAGGCACTTCAAGGCGCAGAGATGATTTCCCGCTGCGATCCT 600
QY 666 GGTCCCAACCCAGTTTGTGTGTCATCATCGGAAAGAGGCGCTTGACCAATAAGAACAT 725
Db 601 GTTCCCAACCCAGTTTGTGTGTCATCATCGGAAAGAGGCGCTTGACCAATAAGAACAT 660
QY 726 CACTAAGCAGACCCAGTCCGCGGATGATTCATATGAAAGAGAACTCTGAGCTGCGAGA 785
Db 661 CACTAAGCAGACCCAGTCCGCGGATGATTCATATGAAAGAGAACTCTGAGCTGCGAGA 720
QY 786 GAAAGCTGTCACCATTCATGCAACCCAGAGGGAATTTGAAAGCATGCGGATGATTC 845
Db 721 GAAAGCTGTCACCATTCATGCAACCCAGAGGGAATTTGAAAGCATGCGGATGATTC 780
QY 846 TGAATATGTCAGAAAGAGGAGATGAGACCAACTAGCCGAGAGATTCCTGAAAT 905
Db 781 TGAATATGTCAGAAAGAGGAGATGAGACCAACTAGCCGAGAGATTCCTGAAAT 840
QY 906 CTGAGCACAATGCTTGTGTGGAAGACTGATTTGAAAGAGGAGAAATTTGAAAG 965
Db 841 CTGAGCACAATGCTTGTGTGGAAGACTGATTTGAAAGAGGAGAAATTTGAAAG 900
QY 966 AATTGAACATGAAAGAGGAGCAAGATTAACAATCTCATCTTTCAGAGATTTGAGATATA 1025
Db 901 AATTGAACATGAAAGAGGAGCAAGATTAACAATCTCATCTTTCAGAGATTTGAGATATA 960
QY 1026 CAACCCGAGAAAGACATCATCTGTGAAGGCAAGTTGAGGCGCTGTGCAAGTGTAGAT 1085
Db 961 CAACCCGAGAAAGACATCATCTGTGAAGGCAAGTTGAGGCGCTGTGCAAGTGTAGAT 1020
QY 1086 AGAGATTATGAGAGAGTGTGAGGCTTTGAAATATGATATGCTGTGTTAACCCCA 1145
Db 1021 AGAGATTATGAGAGAGTGTGAGGCTTTGAAATATGATATGCTGTGTTAACCCCA 1080
QY 1146 CTCCGATATCTTCCAGCTGTGATACCCCATCAACAGTTTGGCCGCTTCCCGATATCA 1205
Db 1081 CTCCGATATCTTCCAGCTGTGATACCCCATCAACAGTTTGGCCGCTTCCCGATATCA 1140
QY 1206 CTCTTATCCAGAGAGATGTGATCTTTCATCCCAACCCAGGCTGTGAGCGGCAT 1265
Db 1141 CTCTTATCCAGAGAGATGTGATCTTTCATCCCAACCCAGGCTGTGAGCGGCAT 1200
QY 1266 CATCGGAGAGAGGAGGCAACATCAAAAGCTGTGAGAGATGTGAGAGATGTGAGAG 1325
Db 1201 CATCGGAGAGAGAGGAGGCAACATCAAAAGCTGTGAGAGATGTGAGAGATGTGAGAG 1260
QY 1326 GATTGCCCCGCGAGAGGCGCAGAGCTCAAGGAAAGATGTATCATCAACCGGAGCACC 1385
Db 1261 GATTGCCCCGCGAGAGTGCCTTCTGCTCAGCGGAAAGATCATCATCACTGGCCACC 1317
QY 1386 GGAAGCCCAATTCAGAGCCAGAGGCGAGATCTTTGGGAAACTGAAAGAGGAAATCTTCTT 1445

Db 1318 GGAATCCCAATTCAGAGCCCAAGGAGCGAGATCTTTGGGAAACTGAAAGAGAAATCTTTT 1377
QY 1446 TAACCCCAAGAGAGAGAGGAGGCTGGAAGCGCATATCAGAGTGCCTTCCACAGCTGG 1505
Db 1378 TAACCCCAAGAGAGAGGAGGCTGGAAGCGCATATCAGAGTGCCTTCCACAGCTGG 1437
QY 1506 CCGGAGATTTGGCAAAAGGTGCGAAGCCGTGAACGAATGCAAACTTAAACAGTGCAGA 1565
Db 1438 CCGGAGATTTGGCAAAAGGTGCGAAGCCGTGAACGAATGCAAACTTAAACAGTGCAGA 1497
QY 1566 AGTCACTGTGCTCGTGAACCAAGCCGATGAAATGAGAAAGATGCTGCAATAT 1625
Db 1498 AGTCACTGTGCTCGTGAACCAAGCCGATGAAATGAGAAATGATGCTGCAATAT 1557
QY 1626 CCGGCACTTCTTGTATGACAGACTGCAAGCGCAAGATGAGGAAATTTGACACAGGT 1685
Db 1558 CCGGCACTTCTTGTATGACAGACTGCAAGCGCAAGATGAGGAAATTTGACACAGGT 1617
QY 1686 GAAAGCAGAGAGAGCAAAATATCCTCAAGGAGTGCCTTCAACAGCGCAGTATGAGCTC 1745
Db 1618 GAAAGCAGAGAGAGCAAAATATCCTCAAGGAGTGCCTTCAACAGCGCAGTATGAGATC 1677
QY 1746 CCACAGCAGACAGCAAAACCAACGATGAT 1775
Db 1678 CCACAGCAGACAGCAAAACCAACGATGAT 1707

RESULT 9
US-09-764-864-749
; Sequence 749, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; Prior application data removed - consult PAM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 749
; LENGTH: 1186
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (511)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (774)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (777)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-749

Query Match 33.4%; Score 1096.8; DB 9; Length 1186;
Best Local Similarity 98.4%; Pred. No. 1.7e-289;
Matches 1145; Conservative 3; Mismatches 12; Indels 4; Gaps 4;

QY 1247 CCCAGGCTGTGGCGCCATATGAGGAGAGGCGCACAATCAAAACAGTGGCGAGAT 1306
Db 25 CCCAGGCTGTGGCGCCATATGAGGAGAGGCGCACAATCAAAACAGTGGCGAGAT 84
QY 1307 TCGCGGAGGCTCTATCAAGATTTGCCCTGCGAAGGCCAGACGTAGGAAAGATGG 1366
Db 85 TCGCGGAGGCTCTATCAAGATTTGCCCTGCGAAGGCCAGACGTAGGAAAGATGG 144
QY 1367 TCATCATCACCGGCGCAACCGAAAGCCAGTTCAAGGCTCAAGGAGCGATCTTTGGAAAC 1426
Db 145 TCATCATCACCGGCGCAACCGAAAGCCAGTTCAAGGCTCAAGGAGCGATCTTTGGAAAC 203
QY 1427 TGAAGAGGAAATCTTTTAAACCCCAAGAAAGAGAGTGAAGCTGGAAGCGATATCAGAG 1486


```
Db      204 TGAAGAGGAAAACTCTTTAAACCCCAAGAAAGTAGAGTGAACCCATTATCAGAG 263
Qy      1487 TGCCCTCTTCCACAGCTGGCGGGGATTTGGCAAGGTGGCAAGACCGTGAACGAACTGC 1546
Db      264 TGCCCTCTTCCACAGCTGGCGGGGATTTGGCAAGGTGGCAAGACCGTGAACGAACTGC 323
Qy      1547 AGAATTAAACAGTGCAGAGTGCATGCTGCTCTGTGACCAAAACGCGAGTGAATAAGAG 1606
Db      324 AGAATTAAACAGTGCAGAGTGCATGCTGCTCTGTGACCAAAACGCGAGTGAATAAGAG 383
Qy      1607 AAGTATGCTGAGAAATTAATCGGAGCTTTCTTTGCTAGCCAGACTGCAAGCGCAAGATCA 1666
Db      384 AAGTATGCTGAGAAATTAATCGGAGCTTTCTTTGCTAGCCAGACTGCAAGCGCAAGATCA 443
Qy      1667 GGGAAATTTG-TACAAACAGTGAAGAGAGAGAGCAAGAAATCCTCAAGGAGTGCCTCA 1725
Db      444 GGGAAATTTGTACAAACAGTGAAGAGAGAGAGCAAGAAATCCTCAAGGAGTGCCTCA 503
Qy      1726 CAGGCGACAGAGTGAAGGCTCCACAGGACAGCAAAACAAAGGATGAATGAGCCCTTC 1785
Db      504 CAGGCGA-NAAGTAGGCTCCACAGGACAGCAAAACAAAGGATGAATGAGCCCTTC 562
Qy      1786 CAACACCTGACAGAAATGAGACCAACGAGCCAGCCAGATCGGAGAGCAAAACCAAGACA 1845
Db      563 CAACACCTGACAGAAATGAGACCAACGAGCCAGCCAGATCGGAGAGCAAAACCAAGACA 622
Qy      1846 TCTGAGGAATGAGAAAGTCTGCGAGAGGCGCCAGAGGACTCTTCCAGAGGCTTGAACCC 1905
Db      623 TCTGAGGAATGAGAAAGTCTGCGAGAGGCGCCAGAGGACTCTTCCAGAGGCTTGAACCC 682
Qy      1906 AGGGGCGGAGAGGAGGCGGAGAGTCAAGCAGGTTTGCGAGAACCAACGAGCCCGCT 1965
Db      683 AGGGGCGGAGAGGAGGCGGAGAGTCAAGCAGGTTTGCGAGAACCAACGAGCCCGCT 742
Qy      1966 CCGGCCCCCAAGGCTTCTGAGGCTTCAAGCAT-CACTTCAACATCACTCGGATCTC 2024
Db      743 CCGGCCCCCAAGGCTTCTGAGGCTTCAAGCATCACTTCAACATCACTCGGATCTC 802
Qy      2025 TCCGAACTCCCAAGAGCTATCCCTTTTATGTTAACTAATAGTGAACGTGTTCAA 2084
Db      803 TCCGAACTCCCAAGAGCTATCCCTTTTATGTTAACTAATAGTGAACGTGTTCAA 862
Qy      2085 GCCAAGCAAAATGACACCCCTTTTCTGTGCAAAATGCTCTGTATAGTGTATATA 2144
Db      863 GCCAAGCAAAATGACACCCCTTTTCTGTGCAAAATGCTCTGTATAGTGTATATA 922
Qy      2145 TTGAAAAGGGAAGATGTTAAGATATGAGCTGTGAGTTACACAGGAGCTGAGCGGT 2204
Db      923 TTGAAAAGGGAAGATGTTAAGATATGAGCTGTGAGTTACACAGGAGCTGAGCGGT 982
Qy      2205 AATATATTTTGAAGAAATATATCAATAATCACTAATCTCAATTTTAAATCAATAT 2264
Db      983 AATATATTTTGAAGAAATATATCAATAATCACTAATCTCAATTTTAAATCAATAT 1042
Qy      2265 TAAATTTTCTTTCTTTTAAAGAAAGAGGCTTTCTAGACTTTTAAATAAATGCT 2324
Db      1043 TAAATTTTCTTTCTTTTAAAGAAAGAGGCTTTCTAGACTTTTAAATAAATGCT 1102
Qy      2325 TTGGAGGTCTCAACGCTGTAGAGAGAGCTTTGAGGCAACCCGACAAATTCACCCAGA 2384
Db      1103 TTGGAGGTCTCAACGCTGTAGAGAGAGCTTTGAGGCAACCCGACAAATTCACCCAGA 1162
Qy      2385 GGGAAATCTGTGGGAAGGACT 2408
Db      1163 GGGAAATCTGTGGGAAGGACT 1186
```

```
APPLICANT: Wang, Tongtong
APPLICANT: Pan, Liqun
APPLICANT: Kalos, Michael D.
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Hosken, Nancy
APPLICANT: Fanger, Gary R.
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Henderson, Robert A.
APPLICANT: McNeill, Patricia D.
APPLICANT: Fanger, Neil
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: AND DIAGNOSIS OF LUNG CANCER
CURRENT FILING DATE: 210121.455C14
CURRENT FILING DATE: US/09/735,705
NUMBER OF SEQ ID NOS: 2000-12-12
SOFTWARE: PatsSeq for Windows Version 3.0
SEQ ID NO 347
LENGTH: 1740
TYPE: DNA
ORGANISM: Homo sapiens
US-09-735-705-347
Query Match 21.2%; Score 697.2; DB 9; Length 1740;
Best Local Similarity 64.9%; Pred. No. 1e-179;
Matches 1116; Conservative 0; Mismatches 538; Indels 66; Gaps 3;
Qy      73 ATGAACAAGCTTTTATCATCGGGAACCTGAGCCCGCCGTCACCGCCGACACTCCGAG 132
Db      1 ATGAACAACCTGATATATGGAACCTGAGAGAGAGCCCGCCCTCGGACATGAAAGT 60
Qy      133 CTCTTTGGGAGACAGAAAGTGCCTGCGGAGACAGTCTGCTGAAATCCGCTACGCC 192
Db      61 ATCTTAAAGACCGCAAGATCCCGTTCGGGACCTTCTGTGAAGACTGCTACGCC 120
Qy      193 TTGTGACTACCCCGACAGAACTGGGCCATCCGCCATCGAACCTCTCGGTAA 252
Db      121 TTGTGACTGCGCCGACAGAGAGTGGCCCTCAAGGCATCGAGGCGCTTCAAGTAA 180
Qy      253 GTGGAATTCATGGGAAATATCATGAAAGTGTATCTCATGCTTAAAGCTTAAGAGC 312
Db      181 ATGAAGTGAACGGAAACCCATGAAAGTGAAGTCACTCGGTCCCAAAAGCAAGATT 240
Qy      313 AGAATAATTCAGATTGGAACATCCCTCTCACTGAGTGGAGAGTGTGAATGACTT 372
Db      241 CGGAACCTTCAGATGAGAAATATCCCGCTCATTTAAGTGGAGTGTGATGTTA 300
Qy      373 TTGGCTCATATGGAACAGTGAAGATGTGAAACAACTCAACACAGCAAGAAACGCC 432
Db      301 CTAGTCCAGATAGAGTGTGAGAGCTGTGACCAAGTGAACCTGACCGAAACTGCA 360
Qy      433 GTTGTGAAGTCAATATGCAACAAAGAAAGCAAAATAGCATGAGAGACTTAAC 492
Db      361 GTTGTAAATGTAACCTATTCAGATGAGAACCAAGCTAGAACACACTAGCAAACTAAT 420
Qy      493 GGGCATCAGTTTGAAGACTACTCTCAAGATTCTTCACTCCGATGAGAGGTGAGC 552
Db      421 GGAATTCAGTTAGAAATTTCACTTGAAGTGAAGCTTATCTCTGATGAAACGCGCC 480
Qy      553 TCCCTTTCGCCCCCTAGAGAGCCAGCGTGGGACCACTTCCCGGAGCAAGGC--- 609
Db      481 CAGCAAAACCCCTTTCAGAGAGCCCGAGAGTCCCGGGGCTTGGGAGAGGGGCTCTCA 540
Qy      610 -----CAGGCCCTGAGGGGCACTTCTCAGGCCACAGATTGAATTCCTGCGGATC 663
Db      541 AGCAGGGGTCTTCAGAGATCCGATCCAGCAAGAAACATGTGATTTGCTTCCGCTG 600
Qy      664 CTGGTCCCAACCAAGTTTGTGTGCTCATCGAAAGAGAGGCTTCAACATAAAGAC 723
Db      601 CTGGTCCCAACCAATTTGTGTGAGCATCATAGGAAGAAAGAGTGCACCATTCGAAAC 660
```

RESULT 10
US-09-735-705-347
; Sequence 347, Application US/09735705
; Patent No. US20020052329A1
; GENERAL INFORMATION:

QY 781 GAGAACCCGTGCACCACTCCATGCGACCCCGAGGGGACTTCTGAAAGCATGCCGCAATGTT 843
 Db 721 GAGAAAGTCGATTACTATTCCTCTCTACTCTCTGAAAGGACCTCTGCGGCTGTGAAGTCATTT 780
 QY 844 CTGGAATCATGCGAGAAAGGCGAGTGTGACCAAACTAGCCGGAAGATTTCTCTGAA 903
 Db 781 CTGGAATTAATGCTATGAGGAAGTCAAGATATPAAATTCACAGAAAGATCTCCCTTGAAG 840
 QY 904 ATCTTGGCACACAAATGGCTTGGTGGAGACATGATTGGAAAAAGAGGAGAAATTTGAAG 963
 Db 841 ATTTTGTGCTCATATAATACCTTTGTGTGAGCTCTTATTTGGTAAGAGAGAAATCTTTAA 900
 QY 964 AAAATTGAACATGAAACAGGAGCAAGATTAACATCTCATCTTTCAGAGATTGAGCAT 1023
 Db 901 AAAATTGACCAAGACACAGACACTAAATATCGATATCTCCATTTGCAAGAAATTGACGCTG 960
 QY 1024 TPCAAACCCGGAAGAAACCATCATCTGTGAAAGGCAACATTTGAGGCTGTGGCCAGTGTGAG 1083
 Db 961 TATTAATCCAGAACGCACTATTACATTTAAAGGCAATGTTGGAACATGTGGCAATGCGCAAG 1020
 QY 1084 ATAGAGATTATGAAAGAGCTGCGTGAAGGCTTTGAAATGATATCTGCTGTTTAACACC 1143
 Db 1021 GAGGAGATCATGAAAGAAATACGGAAGCTTATGAAATGATATGCTTCTATGATCTT 1080
 QY 1144 CACTCGGAT-----ACTTCTCCA 1162
 Db 1081 CAAGCACATTTAATCTCGATTAATCTGAACGCTTGGGCTGTTCGCCACCATTTCA 1140
 QY 1163 GCGCTTACCCCATACCAAGTTTGGCCCGCTTCCGCAATCATCTC----- 1208
 Db 1141 GGGATGCCACTCCCACTCAGGGGCCCTTTCAAGCATGACCTCTCCCTAACCCGACATTT 1200
 QY 1209 --TTATCCAGAGCAGAGATTTGTGAATCTCTTCAATCCCAACCCAGGCTGTGGGGCCATC 1266
 Db 1201 GAGCATATCAGAAACGGAAGACTTTATCTTTTATCCAGCTCTATCACTGTGGGCCATC 1260
 QY 1267 ATCGGAAAGAAAGGGGCAACATCAACAGCTGTGGAGATTTGCGCGAGGCTCTATCAAG 1326
 Db 1261 ATCGGCAAGCAGGGCCAGACATCAAGCAGCTTCTGCTTTGTGTGAGCTTCAATTAAAG 1320
 QY 1327 ATTGCCCTGTGGGAAAGGCCCAAGCTCAAGCAAGATGATGTCAATCATCCGGGCCACCG 1386
 Db 1321 ATTGCTCCAGGGGAAGCACAGATGCTMAAGTGAAGTATGATTAATCACTGACACCA 1380
 QY 1387 GAAGCCCAAGTCCAAAGGCCCAAGGACGATCTTTGGGAAACAGAAAGAGAAATCTTCTT 1446
 Db 1381 GAGGCTCACTTCAAGGCTCAGGGAAGATTTATGAAAAATTTAAAGAGAAATCTTGT 1440
 QY 1447 AACCCCAAGAAAGTGAAGCTGGAAGCGCATATCAAGTGGCCCTTTCCACAGCTGGC 1506
 Db 1441 AGCTCTTAAGAAAGGTGAACCTTGAAGCTCATATCAGAAGTGCATCTTTGTGCTGGCC 1500
 QY 1507 CGGGTGATTTGGCAAAAGTGTGCAGAACCTGTAAACGACTGAGAACTTAACCAATGCAAA 1566
 Db 1501 AGAGTTATTGAAAAAGAGGCAAAACGGTGAATTAACCTTCAGATTTGTCAAGTGCAGAA 1560
 QY 1567 GTCAATCGGCTCGTGAACCAAGCCAGATGAATAAATAGAGAAAGTATCGTCAAGATATTC 1626
 Db 1561 GTTGTGTGCTCTCGTAGCAGACACCTGATAGATATCAAAAGTGTGTCAAAATTAAC 1620
 QY 1627 GGGCACTTCTTTGTAGCAGACATGCAACAGCGCAAGATCAAGGAAATTTGTAACAGAGTG 1686
 Db 1621 GGTCACTTCTATGCTTGCGCAGGTGGCCCAAGAAATTTCAAGAAATTTCTGACTCAGGTA 1680
 QY 1687 AAGCAGCAGAGCAGAAATACCTTCAGGAGTGGCTCAC 1726
 Db 1681 AAGCAGCAGCACAACAGAAAGGCTCTCAAAAGTGCACAC 1720

RESULT 12
 US-09-897-778-347
 ; Sequence 347, Application US/0989778
 ; Patent No. US20020147143A1

```

? GENERAL INFORMATION:
? APPLICANT: Wang, Tonglong
? APPLICANT: Marnetakis, Margarita
? APPLICANT: Fanger, Gary R.
? APPLICANT: Vedwick, Thomas S.
? APPLICANT: Carter, Darick
? APPLICANT: Watanabe, Yoshihiro
? APPLICANT: Henderson, Robert A.
? APPLICANT: Peckham, David W.
? APPLICANT: Fanger, Neil
? TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
? TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
? FILE REFERENCE: 210121.455C16
? CURRENT APPLICATION NUMBER: US/09/897,778
? CURRENT FILING DATE: 2001-06-28
? NUMBER OF SEQ ID NOS: 467
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 347
? LENGTH: 1740
? TYPE: DNA
? ORGANISM: Homo sapiens
US-09-897-778-347

```

Query Match	21.2%;	Score 697.2;	DB 9;	Length 1740;
Best Local Similarity	64.9%;	Pred. No. 1e-179;		
Matches 1116;	Conservative 0;	Mismatches 538;	Indels 66;	Gaps 3
QY	73	ATGACAAGCTTTTACATCGGAGAACCTGAGCCGCCCGCGTCAACCGCGCAACCTCCGCGAG	132	
Db	1	ATGACAAACTGTATATCGGAAACTCTGAGAGAACGCCGCCCTCGGACTTAAAGT	60	
QY	133	CTCTTTGGGACAGAAAGCTGCCCTGACGGGACAGGTCCTGTAACTCCGCTACGCC	192	
Db	61	ATCTTCAAGACCGCAAGATCCCGGTTCGGGACCTCTTCGTGAAGACTGCTACACCG	120	
QY	193	TTGCTGGAACACCCCGACACAGAACTGGGACCATCCGCGCATCGAACCTCTCGGGTAA	252	
Db	121	TTGCTGGAACGCCCGGACGAGAGCTGGGCCCTCAAGGCATCGAGCGCCTTTCAGTTAA	180	
QY	253	GTCGATTTGCAATGGGAAATCATAGAAATTGATTACTCATGTCCTTAAAGCTAAGAGC	312	
Db	181	ATAGAACTGACCGGAAACCATAGAGTTGAGACATCGGTCCAAAAAGCAAGATT	240	
QY	313	AGGAATTTGATTTGGAAATCCCTCCCTCACTGACATGGGAGGTTTGGATGACTT	372	
Db	241	CGGAACTTTAGATGACAAATATCCCGCTCATTTTAAAGTGGAGGTGCTGGATATTTA	300	
QY	373	TTGGCTCAATATGGGACATGAGAAATGTGAAACAAATCAACACAGACACAGAAACGCC	432	
Db	301	CTAGTCAGATATGGAGTGTGGAGAGCTGTGAGAACTGAACATGACTCGGAAACTGCA	360	
QY	433	GTGTGCAAGCTCATATTCGACAAAGAAAGCAAAAAATGGCATGAGAGACTAAGC	492	
Db	361	GTGTGAATATTAACCTATTCAGTAAAGAACAAAGCTAGACAAAGCATAGACAAACTGAT	420	
QY	493	GGGCATCAGTTTGAAGACTACTCCTTCAAGATTTCTACATCCCGATGGAAGAGTGAGC	552	
Db	421	GGATTTCAAGTTAAGAAATTTCACTTGAAGTAGTACTATTTCCCTGTATGAAACGAGGCC	480	
QY	553	TCCCTTTGGCCCCCTGACGAGGCCAAGCGTGGGAGCACTTTTCCCGGAGCAAGGC---	609	
Db	481	CAGCAAAACCCCTTTGACGAGCCCGCAGAGTCGCGGGGGCTTGGGAGAGGGGCTCTTCA	540	
QY	610	-----CAGCCCTCGGGGSCACTTCTCAGGCCAGACAGATTGATTTTCCGCTGGCATC	663	
Db	541	AGGCAAGGGGTCTCCAGGATCCGTATCCAAACAGAAACCAATGATATTTGCTTCGCGCTG	600	
QY	664	CTGTGTCCTCCACCAAGTTTGTGTGCATCATCGAAAGGAGGGCTTGACCATTAAGAC	723	
Db	601	CTGTGTCCTCCACCAATTTGTTGAGGCATCATAGAAAAAAGGTGCAACATTTGGGAC	660	
QY	724	ATCATTAAGCAGACCATGTCGGGGTAGATATCCATAGAAAAAGAACTCTGAGACTGCA	783	

```

Db 661 ATCAACAAGACCCAGCTTAATATGATGTCCACCGTAAAGAAATGCGGGGCTGCT 720
QY 784 GAGAGGCTGTGACATTCATGCCACCCAGAGGGGACTTGTGAAGATGCCGATGTT 843
Db 721 GAGAGGCTGTGACATTCATGCCACCCAGAGGGGACTTGTGAAGATGCCGATGTT 780
QY 844 CTGGAATTCATGAGGAGGAGGAGATGAGCAAACTAGCCGAGAGATCTCTGAAA 903
Db 781 CTGGAATTCATGAGGAGGAGGAGATGAGCAAACTAGCCGAGAGATCTCTGAAA 840
QY 904 ATCTTGGCACAATGCTGTGTTGGAAGACTGATTTGAAAAAGAGCAGAAATTTGAG 963
Db 841 ATTTAGCTCAATTAATTTGTTGGAGCTCTTATTTGTTAAAGAGAAATCTTAAA 900
QY 964 AAAATTGAACTGAAACAGGAGCAAGTAACTATCTATCTTTGAGAGATTTGAGCTA 1023
Db 901 AAAATTGAACTGAAACAGGAGCAAGTAACTATCTATCTTTGAGAGATTTGAGCTG 960
QY 1024 TACAAACCGGAAAGACATCATGTGTAAGGGCAGTTGAGGCTGTGAGCTGAG 1083
Db 961 TATAATCCAGAACGCAATTAATGAGTTAAAGGCAATTTGAGACATGTGCCAAAGCTGAG 1020
QY 1084 ATAGAGATTATGAAAGAGCTGCTGAGGCTTTGAAAAATGATATGCTGCTTTAAACC 1143
Db 1021 GAGGAGATCATGAAAGAAATACAGGAGCTTATGAAATGATATGCTTATGATCTT 1080
QY 1144 CACTCCGAT-----ACTTCTCA 1162
Db 1081 CAAGCACAATTAATTCGTGATTAATCTGAAAGGCTTGGGTCTGTCCACCACTTCA 1140
QY 1163 GCTGTACCCCATCACCAGTTGCGCGTCCCGCATCATCTC----- 1208
Db 1141 GGGATCCACCTCCCACTCAGGGGCCCTTACGACATGACCTCCCTACCCGAGTTT 1200
QY 1209 --TTATCCAGACAGAGATTGTGAATCTTTCATCCCAACCCAGCTGTGGGCGCATC 1266
Db 1201 GAGCATACTGAAAGAGAGCTGTTCATCTGTTATCCGAGCTTATCACTGCTGCTCATC 1260
QY 1267 ATCGGAGAGAGGGGACACATCAACAGCTGGGAGATTGGCCGAGGCTTATCAAG 1326
Db 1261 ATCGGAGAGAGGGGACACATCAACAGCTGGGAGATTGGCCGAGGCTTATCAAG 1320
QY 1327 ATTTGCCCCGAGAGGAGGAGGCTCAGCGTCAAGAAAGGATGTCATCATCAACCGGACACCG 1386
Db 1321 ATTTGCTCCAGCGAGACACATGCTTAAAGTGAAGATGCTTATCATCTGGAACACCA 1380
QY 1387 GAAGCCAGTTCAAGGCCAGGAGCGATCTTTGGGAAACTGAAAGAGGAAATCTTTT 1446
Db 1381 GAGGCTCACTTCAAGGCTCAGGGAAGAAATTTATGAAAAATTAAGAAAGAAATCTTTGTT 1440
QY 1447 AACCCCAAGAGAGTGAAGCTGGAAGGCAATTCAGAGTCCCTTCCAGAGTGGC 1506
Db 1441 ACTCCCTAAAGAGAGTGAAGCTTGAAGCTCATATCAGAGTCCCTTGTGCTGGC 1500
QY 1507 CGGGGATTTGGCAAGAGTGGCAAGACGTTGAAGCACTGCAAGACTTAAACAGTGCAGAA 1566
Db 1501 AAGATTATGAAAGAGAGGAGCAAAAGCTGAATGAAGCTTCAAGATTTGCTAAGTGAAGA 1560
QY 1567 GTCATGCTGCTGTGACCAAAAGCCAGATGAAGTGAAGAGATGATGCTCAGATTTATC 1626
Db 1561 GTTGTGTTCCCTGTGACCAAGACCTGATGAAGATGACCAAGTGTGTCAAAAATTAAGT 1620
QY 1627 GGGCACTTTCTTTGCTAGCCAGACTGCAAGCCGCAAGATCAAGGAAATTTGAAACAAGGG 1686
Db 1621 GGTCACTTTCTATGCTTGGCAGGTTGCCAGAGAAAAATTCAGGAAATTCGACTCAGGTA 1680
QY 1687 AAGCAGAGGAGCAGAAATACCTTCAGGAGTGCCTCAG 1726
Db 1681 AAGCAGACCAACACAGAAAGCTTGCAGAAAGTGACAC 1720

```

```

; Sequence 347, Application US/10007700
; Publication No. US20030064947A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Li, Samuel X.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Fanger, Neil
; APPLICANT: Rector, Marc W.
; APPLICANT: Durham, Margarita
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Peckman, David W.
; APPLICANT: Cai, Feng
; APPLICANT: Foy, Teresa M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.45C17
; CURRENT FILING DATE: US/10/007,700
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 347
; LENGTH: 1740
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-007-700-347

```

```

Query Match      21.2%  Score 697.2; DB 13; Length 1740;
Best Local Similarity 64.9%; Pred. No. 1e-179;
Matches 1116; Conservative 0; Mismatches 538; Indels 66; Gaps 3;

QY 73 ATGAACAAGCTTTACATCGGAGAACTGAGCCCGCTGACCCGAGACGACTCCGGAG 132
Db 1 ATGAACAAGCTTATATGAGAACTCAGCGAGAAAGCGCCCTCGGACCTAGAAAGT 60
QY 133 CTCTTTGGGAGACAGAAAGCTGCCCCCTGGCGGAGACAGTCTGTAAGTCCGGCTAAGCC 192
Db 61 ATCTTCAAGAGAGCCAGAAATCCGGGTGAGGAGCCCTTCTGTGAAGCTGAGCTACGGC 120
QY 193 TTGCTGAGTACCCGACAGAACTGGGCTATCCGGCCATCGGACATGAGAACCTCTCGGCTAAA 252
Db 121 TTGCTGAGTACCCGACAGAACTGGGCTATCCGGCCATCGGACATGAGAACCTCTCGGCTAAA 180
QY 253 GTGGAATTCATGAGGAAATCATGAAAGTGAATTAATCTAGTCTTAAAGACTAAGAGAGC 312
Db 181 ATGAAATCTCAGCGGAAACCCATAGAACTTGAGCACTCGATCCCAAAAGGCAAAAGGATT 240
QY 313 AGGAAATTCATGATTTGAAATATCCCTCTCTACCTGAGTGGGAGGTTGTGATGACTT 372
Db 241 CGGAAATTCATGATTTGAAATATCCCTCTCTATTTACATGAGGAGGTTGTGATGACTT 300
QY 373 TTGGCTCAATATGAGGACATGAGAAATGTGAAACAGTCAACACAGACACAGAAACGCC 432
Db 301 CTAGTCAATATGAGAGTGTGAGAGCTGTGAGAAAGTGAACATGACATCTCGGAAATGCA 360
QY 433 GTTGTCAACGTCAATATGCAACAGAGAAAGAAAGAAATTAAGCCATGAGAAAGCTAAGC 492
Db 361 GTTGTAAATGTAACCTATTCAGTAAAGGACAAAGCTGACAAAGCACTAAGCAAACTGAAT 420
QY 493 GGGCATCAGTTTGAAGAACTATCCCTTCAAGATTTCTCAATCCCGGATGAAGAGTGAAGC 552
Db 421 GGAATTCAGTTAAGAAATTCACCTTGAAGAGTACCTTATATCCCTTAATGAAGAGGCCGCC 480
QY 553 TCCCTTGGCCCCCTCAGCGAGCCAGCGTGGGAGCACTTTCCCGGAGCAAGAGC--- 609
Db 481 CAGCAAAACCCCTTGAAGAGCCCGAGAGTCCCGGGGCTTGGGAGCAAGAGGCTCTCTCA 540

```

QY 610 -----CAGGCCCTGGGGGCACTTCTCAGGCCAGACAGATGATTTCCCGCTGGGATC 663
DB 541 AGGCAAGGGGTCTCCAGGATCCGTATCCAGAGAGAAACATGATGATTTCCCTGGGCTG 600
QY 664 CTGGTCCCAACCACTTTGTTGGTCCATCATCGGAAAGAGGGCTTGACATTAAGAC 723
DB 601 CTGGTCCCAACCACTTTGTTGGAGCCATCATAGGAAAGAGGGCCACCATTCGGAAC 660
QY 724 ATCACTAGCAGACCCAGTCCCGGGTAGATATCCATTAAGAAAGAACTCTGGAGCTGCA 783
DB 661 ATCACTAGCAGACCCAGTCTTAATATGATGATTCACCTTAAGAAAGAGGGGGCTGCT 720
QY 784 GAGAGCCTGTACCATTCATGCAACCCAGAGGGGACTTCTGAAGATGCGGATGAT 843
DB 721 GAGAGGTGATTAATCTCTCTACTCTGAGGACCTCTGGGGCTTGAATGATTT 780
QY 844 CTGGAATCATGCGAAGAGGCGAGATGAGCAAACTAGCCGGAAGATTTCTTGAAA 903
DB 781 CTGGAAGATTAAGCAATTAAGAGCTCAAGATTAATAATTCAGAGAGATCCCTTGAG 840
QY 904 ATCTTGACACAAATGGCTTGGAAGTGAATGGAAGTGAAGAAAGAGAAATTTGAG 963
DB 841 ATTTAGCTATAATTAATCTTTGGAGCTTATGGAAGAGAGAAATCTTAAA 900
QY 964 AAAATGAAATGAAAGAGGAGCCAGATTAACAATCTCATCTTTCAGGATTTGAGATA 1023
DB 901 AAAATGAGAGAGACAGACATAAATACGATATCTCATGTCAGGATTAAGCTG 960
QY 1024 TACACCCCGAAGAACATCATCTGTGAAGGGCACAGTTGAGGCTGTGCAAGTGTAG 1083
DB 961 TATATCCAGAACCACTATTACAGTTAAAGGCAATGTGAGCATGTGCCAAAGCTGAG 1020
QY 1084 ATAGAGATTAAGAGAGCTGCGGCTTGAAGCTTGAAGTATGCTGCTGTTAAAC 1143
DB 1021 GAGAGATCAATGAAGAAATCAAGGAGCTTATGAATATGATGCTTATATAATCTT 1080
QY 1144 CACTCCGAT-----ACTTCTCA 1162
DB 1081 CAGCACAATTAATCTGATTAATCTGAACGCTTGGGTCTGTTCCACCATCTTCA 1140
QY 1163 GCTGTACCCCATCAACCAATTTGGCCGTTCCGCAATCACTC-----1208
DB 1141 GGGATGCACTCCCACTCAGGGCCCTTCAAGCCATGACTCTCCCTAACCCGAGATT 1200
QY 1209 --TTATCAGAGAGAGATTTGAATCTTCAATCCCAACCGAGCTGTGGGCTATC 1266
DB 1201 GAGCAATCAAGAAAGAGAGCTGATCTGTTATCCAGCTTATCAAGTGGGCTATC 1260
QY 1267 ATCGGAAAGAGGGGAGACATCAACAGCTGGGAGATTTGCGCGAGGCTTATCAAG 1326
DB 1261 ATCGGAGAGAGGGCCAGACATCAACAGCTTTCTGCTTGTGAGAGCTTCAATTAAG 1320
QY 1327 ATTGCCCCCTCGAGAGGCCAGAGCTGAGGAAAGATGTCATCACCGGCGCACG 1386
DB 1321 ATTGCTCCAGCGGAGAGCAGAGTCTAAAGTGAAGTGTATATCACTGAGCACCA 1380
QY 1387 GAAGCCAGTTCAAGGCCAGGAGCGATCTTTGGGAACTGAAAGGAGAAATTTCTT 1446
DB 1381 GAAGCTCAAGTTCAAGGCTCAGGAGAAATTTATGAGAAATTAAGAAAGAAATTTGTT 1440
QY 1447 AACCCAAAGAGAGTGAAGTGAAGCGCATATCAGATGCTCTTCCAGAGCTGCG 1506
DB 1441 AGTCTAAAGAGAGAGTGAAGTGAAGCTCATATAGATGCTCTTGTGCTGAGC 1500
QY 1507 CGGGTATTTGGCAAGGTGCAAGCCGTGAAGCAACTGAGAACTTTAACAGTGCAGAA 1566
DB 1501 AGAGTTATTTGGAAGAGGAGCAAGCGTGAATGAATCTTCAAGATGCGAGAA 1560
QY 1567 GTCATGTGCTGTGAGCAAGCGCATGAGAAATGAGAAATGATCTGCAAGATTAATC 1626
DB 1561 GTTGTGTCTCTGTGAGCAAGCCTGATGAGATGACCAAGGTGTTGCAAAATTAATC 1620
QY 1627 GGGCACTTTCTTGTGAGCAAGCTGCAAGCGCAAGATCAGGAAATTTGTAACAGGTG 1686

DB 1621 GTTCACTTTATGCTTGGCAGGTTGCCAGGAAATTTCTGAGAAATTTCTGATCAAGTA 1680
QY 1687 AAGCAGAGAGAGCAAAATACCTTCAGGAGTCCGCTCAC 1726
DB 1681 AAGCAGAGCAACAAAGAGGCTCTGCAAGTGCACAC 1720

RESULT 14
US-10-117-982-347
; Sequence 347, Application US/10117982
; Publication No. US20030138438A1
; GENERAL INFORMATION:
; APPLICANT: Foy, Teresa M.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kaios, Michael D.
; APPLICANT: Mericle, Barbara
; APPLICANT: Spies, Gregory A.
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Tonglong
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C18
; CURRENT APPLICATION NUMBER: US/10/117,982
; CURRENT FILING DATE: 2002-04-05
; NUMBER OF SEQ ID NOS: 484
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 347
; LENGTH: 1740
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-117-982-347

Query Match 21.2%; Score 697.2; DB 15; Length 1740;
Best Local Similarity 64.9%; Pred. No. 1e-179;
Matches 116; Conservative 0; Mismatches 538; Indels 66; Gaps 3;

QY 73 ATGAAACAAGCTTTACATCGGAAACCTGAGCCCGCTGACCGCGAGACCTCCGCGAG 132
DB 1 ATGAAACAAGCTTTATCGGAAACCTGAGAGAGCGCCCTCGGACCTTGAAGAAAGT 60
QY 133 CTCTTGGGAGCAGAAAGCTGCCCCCTGGCGGACAGTCTCTGTAAGTCCGAGTACGCC 192
DB 61 ATCTTGAAGAGCGCAAGATCCCGGTGTGGAGCCCTCTGTGTAAGATCGGCTACGCG 120
QY 193 TTGTGTGACTTACCCGACCGAAACTGGGCTATCGGCGCATCGAGACCTCTCGGTTAAA 252
DB 121 TTGTGTGACTGCCCCGAGCGAGAGCTGGCCCTCAAGGCCATCGAGGCGCTTTCAAGTAAA 180
QY 253 GTGAAATTTGATGAGGAAATCATGAAATGTTACTCACTCTCTTAAAGGCTAAGAGC 312
DB 181 ATGAAACTGACCGGAAACCATTAAGATTTGAAGCTCGGCTCCAAAAAGCAAGGATTT 240
QY 313 AGGAAATTTCAATTTGAGAAATCTCTCTCACTGAGTGGAGGATTTGATGAGACTT 372
DB 241 CGGAAATTTCAATTTGAGAAATCTCTCTCACTGAGTGGAGGATTTGATGAGACTT 300
QY 373 TTGGCTCATATGAGAGAGTGGAAATGTTGAAACAACTCAACACAGACAGAAACCGCC 432
DB 301 CTAGTCCAGTATGAGATGTTGAGAGAGCTGTGACCAAGTGAACCTGACGAAATCGCA 360
QY 433 GTTGTCAAGCTCATATGCAACAGAGAAAGCAAAATTAAGCAATGAGAGGCTAAGC 492
DB 361 GTTGTAAATGTAATCTATTTCAATAGAGCAAGAGCTGAGAAAGCACTAAGCTAAT 420
QY 493 GGGCATCAAGTTTGAAGACTATCTCTTCAAGATTTCTCAATCCGAGTGAAGAGTGAAGC 552
DB 421 GGAATTCAGTTAGAGATTTCACTTGAAGATGCTTATATCTCTGATGAAAGGCGCGCC 480

```
QY 553 TCCTCTTCCGCCCCCTCAGCGAGCCCAAGCTGGGAGCACTCTTCCCGGAGCAAGGC--- 609
Db 481 CAGCAAAACCCCTTGGAGAGCCCGAGGTCGCGGGGGCTTGGGCAAGGGGCTCTCA 540
QY 610 -----CAGCGCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTTTCCGCTGGGATC 663
Db 541 AGGCAAGGGGCTCTCAGGATCCGTATCCAGCAAAACCATGATGATTTCTCTGGGCTG 600
QY 664 CTGGTCCCAACCGATTGTTGGTGCCATCATCGGAAAGAGGGCTTGACCTAAGAAC 723
Db 601 CTGGTCCCAACCGATTGTTGGAGCCATAGGAAAGAGGGCCACCATTCGGAAC 660
QY 724 ATCACTAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAACTCTGAGCTGCA 783
Db 661 ATCACTAGCAGACCCAGTCTTAAATCAATGTCACCTGAAAGAAAATCGCGGGGCTCT 720
QY 784 GAGAGCCTGTACACCATCCATGCCACCCCAAGGGGAGCTTGAAGCATGCCGATGAT 843
Db 721 GAGAAAGTATTAATCATCTCTCTACTCTGAAAGGCACTCTGGCGCTTGTAAGCTAT 780
QY 844 CTTGAATCATGCAAGAGGAGGAGATGAGACCAACTAGCCGAGAGATTCCTGAAA 903
Db 781 CTGGAGATTATGCAATAGAGAGCTCAAGATATTAATCAAGAAAGATCCCTTGAG 840
QY 904 ATCTTGGACAACAATGGCTTGGTGGAAAGCTGATTGGAAGAAAGAGCAAAATTTGAG 963
Db 841 ATTTAGCTCATATAAATCTTTTGGAGCTCTTATTTGGTAAAGAAAGAAATCTTAAA 900
QY 964 AAAATTTGAACATGAACAGAGCAACAAGTAACAACTCATCTTTGAGAGATTGAGATA 1023
Db 901 AAAATTTGAGCAAGACACAGACCTAAATACAGATATCTCCATTCAGGAAATTTGAGCTG 960
QY 1024 TACAAACCCGGAAGAACCATCATCTGTGAAGGCAAGTGGAGCCCTGGGCTGAG 1083
Db 961 TATAATCCAGAACGACATTAACAAGTTAAAGGCAATGTGGAACATGTCCAAAGCTGAG 1020
QY 1084 ATAGAGATTATGAGAGAGCTGCTGAGGCTTTGAAATGATATGCTGGCTTTAAACC 1143
Db 1021 GAGAGATCATATAAGAAATACAGGAGTCTTATGAAATGATATGCTTATGAAATCT 1080
QY 1144 CACTCCGAT-----ACTTCTCCA 1162
Db 1081 CAAGCAATTTAATCTCTGATTTAATCTGAACGCTTGGCTGTCTTCCCACTTCA 1140
QY 1163 GCCTGATCCCATCATCCAGTTTGGCCCTTCCCGATCATCACTC----- 1208
Db 1141 GGGATGCCACTTCCACTCAGGGGCCCTTCAAGCCATATCTCTCCTAACCCGAGTTT 1200
QY 1209 --TTATCCAGAGAGAGATGTGAATCTTTCATCTCCAAACCAAGGCTGTGGGCGCATC 1266
Db 1201 GAGCAATCGAAGACGAGACTGTTCATCTGTTTATCCAGCTCTATCAATCGGTGCATC 1260
QY 1267 ATCGGAGAGAGGGGACACATCAAAACGCTGGCGAGATTGGCGGAGCTCTATCANG 1326
Db 1261 ATCGGAGAGAGGGGACACATCAAGCACTTCTCGCTTGTGAGACTTCAATTAAG 1320
QY 1327 ATTCGCCCCCGGAGAGCCCAACGCTCAGCGAAAGGATGGTCACTCAACCGGCGCAACG 1386
Db 1321 ATTCGCCCCCGGAGAGCCCAAGTGTAAAGTGAAGTGTGTATCACTGGACACA 1380
QY 1387 GAGAGCCAGTTCAGAGCCGAGGACGATCTTTGGGAACTGAAAGAGGAAACTTCTTT 1446
Db 1381 GAGGCTCACTTCAAGCTCAGGGAAGAAATTTATGAAAAATTTAAAGAAACTTGT 1440
QY 1447 AACCCCAAGAAAGTGAAGCTGAAGCGCATATCAAGTGCCTCTTCCACAGCTGGC 1506
Db 1441 AGTCCCTAAAGAGAGGTGAACCTTGAAGCTCATATCAGAGTGCATCTTGTCTGGC 1500
QY 1507 CGGCTGATTTGCAAGAGTGGCAAGACCTGGAAGAACTGAGAACTTAAACCAATGAGAA 1566
Db 1501 AAGATTTATGAAAAAGAGCAAAACGTTGAATGAACCTTCAAGATTTGTCAATGAGAA 1560
QY 1567 GTCATGTGCTGTGACAAACGCGAGATGAATAAGAGAAAGTGAATCGTCAGAAATTATC 1626
```

```
Db 1561 GTTGTGCTCTCTGACACAGACCTGATGAGATGACCAAGTGTGTGCAAAATAACT 1620
QY 1627 GGGCACTTCTTTGTAGCCACATGCAACAGCGCAAGATCAAGGAATGTCAACAGGTG 1686
Db 1621 GGTCACTTCAATGCTTGGCAGGTGCGCAGAAAAAATTCAGGAATTTCTGACTCAGGTA 1680
QY 1687 AAGCAGCAGAGCAAAATACCTCAGGAGTGCCTTAC 1726
Db 1681 AAGCAGCAGCAACAACAGAAAGCTCTGCAAAAGTGAACAC 1720
```

RESULT 15

```
US-10-117-982-478
; Sequence 478, Application US/10117982
; Publication No. US20030138438A1
; GENERAL INFORMATION:
; APPLICANT: Foy, Teresa M.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Matanabe, Yoshihiro
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Meticle, Barbara
; APPLICANT: Spies, Gregory A.
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Tonglong
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C18
; CURRENT APPLICATION NUMBER: US/10/117,982
; CURRENT FILING DATE: 2002-04-05
; NUMBER OF SEQ ID NOS: 484
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 478
; LENGTH: 1740
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-117-982-478
```

```
Query Match 21.2%; Score 697.2; DB 15; Length 1740;
Best Local Similarity 64.9%; Pred. No. 16-179; Indels 66; Gaps 3;
Matches 1116; Conservative 0; Mismatches 538;
QY 73 ATGAACAAGCTTTAATCATCGGGAACCTGAGCCCGCGCTCAACGCGCGACGACCTCCGCGAG 132
Db 1 ATGAACAAGCTTTAATCATCGGGAACCTGAGCCCGCGCTCAACGCGCGACGACCTCCGCGAG 60
QY 133 CTCTTGGGAGCAGGAAGCTGCCCCCTGGCGGAGACAGTCTGCTGAAGTCCGCTACGCC 192
Db 61 ATCTTCAAGGAGCGCAAGATCCCGGTGCGGAGACCTTCTGATGAAGACTGGCTACGCG 120
QY 193 TTCTGAGACTACCCCGACCAAACTGGGGCACTCCGCGCATGAGAGCCCTTCGCGGTAAA 252
Db 121 TTCTGAGACTGCGCGACAGAGCTGGGCTTCAAGGCGCATGAGGCGCTTTCAGGTAAA 180
QY 121 TTCTGAGACTGCGCGACAGAGCTGGGCTTCAAGGCGCATGAGGCGCTTTCAGGTAAA 180
QY 253 GTGAATTGCATGGGAAATCATGGAAGTTGATTAATCACTGATCTTAAAAAGCTTAAGAGC 312
Db 181 ATGAACCTCAAGGGAACCACTAGAAAGTTGAGCACTGGGTCCCAAAAAGGCAAGGAT 240
QY 313 AGAAAAATTCAGATTGGAACATCCCTCTCACTGCACTGCAAGTGGAGGTGTTGATGACCT 372
Db 241 CGGAAACTTCAAGTACGAATATCCCGCTCATTTACATGAGGAGGTGCTGATAGTTTA 300
QY 373 TTGGCTCAATATGGAACAGTGAAGATGTGGAACAAGTCAACACAGACAGAAACCGGC 432
Db 301 CTAGTCCAGTATGAGAGTGTGAGAGCTGTGAGCAAGTGAACCTGACTGGAACCTGCA 360
QY 433 GTTGCAAGCTCATATGCAACAAGAAAGCAAAATATAGCCATGAGAAAGCTAAGC 492
Db 361 GTTGAATATTAATCTATTCAGTAAAGACCAAGCTTACCAACACTAAGCAACTGAAT 420
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2004, 06:21:04 ; Search time 8204.67 Seconds

(without alignments)
11948.985 Million cell updates/sec

Title: US-09-270-437D-8
3283

Perfect score: 1 ggcgcgagagagcgcgagga.....aaccttgaaatgttatttc 3283

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_hiv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1935.2	58.9	3592	11	AK049196 Mus muscu
2	1765.4	53.8	3262	11	AK077404 Mus muscu
3	1622	49.4	2364	11	AK044984 Mus muscu
4	941.6	28.7	999	12	BMS61242 AGENCOURT

5	842.2	25.7	874	13	BO691878
6	830	25.3	851	13	BQ221568
7	809	24.6	832	13	BX644668
8	799.2	24.3	836	13	BQ212500
9	797	24.3	1089	12	BM806132
10	759.8	23.1	894	13	BUI56245
11	752.6	22.9	781	12	BG575889
12	751.4	22.9	940	14	CD243351
13	749.8	22.8	824	14	CF593505
14	746	22.7	889	12	BG748346
15	718.8	21.9	762	13	BU601634
16	695	21.2	2202	11	AK011689
17	695	21.2	4100	11	AK088465
18	691.8	21.1	939	13	BQ961314
19	689.8	21.0	2780	11	AK013940
20	689	20.9	724	12	BM977219
21	686.6	20.9	2718	11	AK044850
22	686.4	20.9	891	13	BUI52527
23	686.4	20.9	1212	13	BU902212
24	677.8	20.6	918	13	BQ212353
25	652.6	19.9	815	12	BI688088
26	651.6	19.8	905	12	BM453327
27	634	19.3	647	13	BU674692
28	631.6	19.2	729	10	BE278075
29	607.2	18.5	1368	12	BG107753
30	606.2	18.5	759	10	BE622021
31	605.4	18.4	1024	9	AL549055
32	595	18.1	628	12	BI682502
33	594	18.1	607	12	BM724394
34	591.8	18.0	759	14	CA320888
35	587	17.9	612	13	BU679082
36	584	17.8	584	13	BX642777
37	571.4	17.4	616	14	CA430674
38	565	17.2	638	12	BM843467
39	561.2	17.1	699	13	BU705477
40	553.8	16.9	1078	12	BM620651
41	548	16.7	594	12	BI585880
42	546.8	16.7	668	13	BY734118
43	542.6	16.5	845	13	BU155790
44	541.6	16.5	707	13	BY720350
45	540.4	16.5	1378	11	AK031773

ALIGNMENTS

RESULT 1	AK049196	3592 bp	mRNA	linear	HTC 20-SEP-2003
LOCUS	AK049196				
DEFINITION	Mus musculus ES cells cDNA, RIKEN full-length enriched library, clone: C330012H03 product: HEPATOCYLLULAR CARCINOMA AUTOMATIGEN homolog [Homo sapiens], full insert sequence.				
ACCESSION	AK049196.1	GI:26093366			
VERSION	AK049196				
KEYWORDS	HTC, CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Carninci, P. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	2				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20499374				
PUBMED	11042159				

REFERENCE
AUTHORS
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carrinci, P., Komoto, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsumoto, Y., Okazaki, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
Riken integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL
MEDLINE
PUBMED
20530913
11076681

REFERENCE
AUTHORS
4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

JOURNAL
REFERENCE
AUTHORS
5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL
REFERENCE
AUTHORS
6 (bases 1 to 3592)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carinci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komoto, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, Y., Tanaka, T., Tomaru, A., Toy, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE
JOURNAL
Direct Submision
Submitted (16-JUL-2001) Yoshinide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.

FEATURES
source
1. 3592
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB: C330012H03"
/db_xref="MGI: 2416441"
/db_xref="taxon: 10090"
/clone="C330012H03"
/clone="C330012H03"
/cell_type="ES cells"
/clone_1ib="RIKEN full-length enriched mouse cDNA library"
34..1811
/note="HEPATOCELLULAR CARCINOMA AUTOANTIGEN homolog [Homo sapiens] (SPTRI|O9YCM1, evidence: FASTY, 88.3%id, 100%length, match=1775)
putative"

ORIGIN
Query Match 58.9%; Score 1935.2; DB 11; Length 3592;
Best Local Similarity 79.7%; Pred. No. 0;

Matches 2720; Conservative 0; Mismatches 469; Indels 224; Gaps 28;
QY 38 GGGGAGCCCGCGGCTCTCGGGAGAGAGAGCATGATGAAACAAGCTTATCATCGGAGAAC 97
3 GGGGAGCCCGCGGCTCTC-GGGAGAGCGGAGATGATGAAACAAGCTTATCATCGGAGAAC 61
QY 98 TGAAGCCCGCGGCTCTCGGAGAGAGAGAGCATGATGAAACAAGCTTATCATCGGAGAAC 157
62 TGAAGCCCGCGGCTCTCGGAGAGAGAGAGCATGATGAAACAAGCTTATCATCGGAGAAC 121
QY 158 TGGCGGAG 217
122 TGGCGGAG 181
QY 218 GGGCATTCGCGGAG 277
182 GGGCATTCGCGGAG 241
QY 278 AAGTTGATTAATCACT 337
242 AAGTTGATTAATCACT 301
QY 338 CT 397
302 CGCT 361
QY 398 ATGTGGAACAAGTCAACAA 457
362 ACGTGAAGCAAGTCAACAA 421
QY 458 GAG 517
422 GAG 481
QY 518 TCAAGATTCT 577
482 TCAAGATTCT 536
QY 578 AGCGTGAGGAGCACT 637
537 -----TGCCTGGAG 580
QY 638 GACAGATTGATTCT 697
581 GACAGATTGATTCT 640
QY 638 GAAAG 757
641 GAAAG 700
QY 758 ATGAG 817
701 ACGAG 760
QY 818 GGAATTCTGAG 877
761 GGAATTCTGAG 820
QY 878 AACTGAG 937
821 AACTGAG 880
QY 938 TTGGAG 997
881 TTGGAG 940
QY 998 TCTCATCTTTGAG 1057
941 TCTCATCTTTGAG 1000
QY 1058 CAGTTGAG 1117
1001 CAGTTGAG 1060

```
QY 1118 AAAATGATATGTGCTGTTAAC----- 1140
Db 1061 AGAAGCATGTGCTGCCGTTAAACAAGCAATCTGATCCAGGGCTAAACCTCAGTG 1120
QY 1141 ----- 1140
Db 1121 CACTTGGCATCTTTTGAATGAGACTGTCTGTGCTTCTTCAAGCAGGGCCCCGNGAGT 1180
QY 1141 -----ACCCACTCCGAGATCTTCTCAAGCTCTGA 1169
Db 1181 TCCCCCAGTCTCTCCCTATCAACCCCTTTGCTACCCACTCCGAGATCTTCTCAAGCTCTGA 1240
QY 1170 CCCCCATCACCAGTTTGGCCCGCTTCCGSCATCATCACTTATCCAGAGCAGAGATTGT 1229
Db 1241 CCTCATACCATTTTGGGCCATTCCCATCATCATCTCTTACCCAGAGCAGAGAGACTGT 1300
QY 1230 GAATCTCTTCAATCCCAACCCAGGCTGTGGGGCCATCATCCGGAAGAGAGGGGCAACAT 1289
Db 1301 AAGTCTCTTCAATCCCAACCCAGGCTGTGGGTGTATCATCGGGAAGAGGGGCAACAT 1360
QY 1290 CAAACACTGGCGAGATTTGCGCGAGGCTCTATCAAGTTTCCCTGCGAGAGCCCA 1349
Db 1361 CAACAGCTGCTGATTTTGTGTGTCTTCATCAAGATTGCTCCAGAGAGGTCCAGA 1420
QY 1350 CGTCAGCGAAAGATGTGATCATCACCGGGCCACCGGAAGCCAGTTCAAGGCCAGGG 1409
Db 1421 TGTCAGTGAAGATGTGATCATCATCTGTCTCTCGAAGCCCAAGTTTAAGGCTCAGGG 1480
QY 1410 ACCGATCTTTGGGAACTGAAAGAGAAAACTTTTAAACCCCAAGAGAGATGAAGT 1469
Db 1481 ACCGATCTTTGGGAAAGTGAAGAGAAAACTTTTAAATCCCAAGAGAGATGAAGT 1540
QY 1470 GGAAGGCGCATATAGAATGCCCTCTTCCAAGCTGGCCGGGTGATTGGCAAGGTGGCAA 1529
Db 1541 GGAAGGCCCATATCCGATGCCCTCTCTGACCGCTGCGGTGAATTGGCAAGGGCGGAA 1600
QY 1530 GACCGTGAAGCAACTGAGCAAACTTAAACAGTGCAGAGTATGCTGTGACCAAG 1589
Db 1601 AACGTGAAGAGGTGAGAACTTGAAGATGCAAGAGTTATGCTGTGACCAAG 1660
QY 1590 GCCAGATGAAGATGAGAGATGATGTCAGAAATTATCGGCACTTTTGTAGCCAGAC 1649
Db 1661 GCCAGAGAAATGAGAGATGATGTCAGAAATTATCGGCACTTTTGTAGCCAGAC 1720
QY 1650 TGCAAGCGCAAGATGAGGAAATTGTAAACAAGGTGAACAGACAGAGCAAAATACC 1709
Db 1721 TGCAACAAGCAAGATGAGGAAATTGTAAACAAGGTGAACAGAGCAAGATACC 1780
QY 1710 TCAGGAGTCTGCTCAAGCGCAGCAAGTGAAGGCTCCACAGGCAACAGCAAAACAG 1769
Db 1781 TCAGGAGTCTGCTCCACAGCGCAGCAAGTGAAGGCTCCACAG--CACACAGCAAGCAACG 1838
QY 1770 ATGAATGTAGCCCTTCCAAACCTGACAGATGAGCAAAACGACAGCAGATCGG 1829
Db 1839 ATGAATGTAGCCCTCCCAACCTGACAG--ATGAGCAAAAC--AGCCAGCAGATCGG 1894
QY 1830 AGCAAAACCAAGACAT--CTGAGGAATGAGAATCTGCGGAGCGCGCAGGGACTGTGCC 1888
Db 1895 AGCAAAACCAAGAGCATCCGGAAGATGCGCAGATCTGCAAGCAGCGAGGCGCTGCG-- 1950
QY 1889 GAGGCCCTGAGAAACCCAGGGGCGGAGAGGGGCGGGAAGTCAAGCAAGTTTGCACA 1948
Db 1951 -AGACTCTACACATCTTGGGATCCAGAGGGGCAAGGGAG--GCCAGTTTGTCCAG 2006
QY 1949 ACCACCG--AGCCCGCTTCGCCGCCCAAGGGCTTCTGAGGCTT--CAGCCATCCACT 2004
Db 2007 AACACCGCTTGGCTGCCGCCCAAGCTTCCCTGCTTCTGAGGCAATACAGGCATCACT 2066
QY 2005 TCACCATCAGCTGGAATCTCTCTGAATCCCAAGAGCGCTATCCCTTTAGTTGAATGA 2064
Db 2067 GCCCATCAACTGAGAT--TCTCTCAAGTTCCCAAGAGCGCTATCCCTTGGTTGAATGA 2125

QY 2065 CATAGTGAAGCTGTTCAAAGCCAAAGCAAAATGCAACACC---TTTTCTGTGCAAAAT 2120
Db 2126 CATAGTGAAGCTGTTCAAAGCCAAAGCAAAATTCCTACCCGTTTCTTGTGTGGAAGT 2185
QY 2121 CGTCTCTGTACATGTGTGTATCATATTTAGAAAGGGA--GATGTTAAGATGTGGCTGTG 2179
Db 2186 TGTCTCTGTACATGTATGTATCATATTCAGAGGGGAAGGATTTAGAAATGTGGCTGTG 2245
QY 2180 GGTTCACAGGGTCTCTGACGGGTATATATTTTAGAATATATATCAATTAATCA 2239
Db 2246 GGTTCACAGGGTCTCTGACGGGTATATATTTTAGAATATATATCAATTAATCA 2305
QY 2240 CTAATCCAAATTTTAAATTCATATTAATTTTCTTTTCTTTTAAAGAAAGCAGCTT 2299
Db 2306 CTAACCAAAATTTTAACTCAAGTGAATTAATTTTCTTTTCTTTTAAAGAAAGCAGCTT 2365
QY 2300 TTCTAGACTTTAAAGATTAAGCTTTGG--GAGGTCTAGGGGTAGAGAGAGCTTTGA 2358
Db 2366 TTCTAGACTTTCAAGAAATTAAGTATTTGGGTAGGTCACCTGTATAGAAAGAACTTTGA 2425
QY 2359 GGC-CACCCGCAAAAATTCACCCAGAGGGAATCTGTCGGAAGGACACTCAGGGCAGT 2417
Db 2426 GCCACCCACAAAATTAATTCAGTTGAGGGAAATCTTGTG--AAGGACCTTAATGGCAGT 2484
QY 2418 TCTGGATCACCTGTGTATGTCAACAGAAAGGATACCGTCTCTTGAAGAGAACTCTGT 2477
Db 2485 TGTGGGTACCTGTGTGTGTCAACAGAAAGGATACCGTCTCTGAGAGGGCCCTAGGCC 2544
QY 2478 CACTCTCATGSCCTGTACTACTCATACCACTTCTCT-----T 2517
Db 2545 CTGGTCACTGTTTCTTCCACACACCAATGCTCAAGTTTCAAGGTTTAACTGA 2604
QY 2518 TGCTTCAAGGTTTAAACTGTTTGTGATCTGCTATATATATCTCTGTCTCTCT 2577
Db 2605 TTTTGGTGTGTTTGTGTTTGTGTTTGTGATCTGTATATATGTCTGTCTCTCC-- 2662
QY 2578 GTTATATCTCCCTCCCTCCCTCCCTCCCTCTTCTCATCTCCATCTTTTGAATTTCT 2637
Db 2663 ----TATCTCCCTCTTAACTTCTCTGTCTTCTTAAACCAAAATCTTTTGAAGTCCCT 2718
QY 2638 CATCCCTCATCTCAATCCCGTATCTAGCAGCC---CGCCGCCCGCAGGCAAAAGCAGT 2694
Db 2719 TCGTTCTTTTCTGTCTCCGTAATCTATGTACCCCTCCCTCCCGCAGGAAAGCAGTG 2778
QY 2695 CTCTGATATCATATCAACACAAAGGAACAAAGCGAACAACACAAACAGCCTCACTT 2754
Db 2779 CTCTGAATATCATATCAACACAAAGGAACAAATGCGAACACACACAGCCTCAGCTT 2838
QY 2755 ACACTTGTATCTCAAAAGGAACAAAGTCAATGTACTTGTCTTACGCTTTGGAAGAG 2814
Db 2839 ACACTGCTTATCTCAAGCGAACAAGAGTCGGCGGTACTTGTCTTACG--ATTGGAAGAG 2897
QY 2815 AAAACAGGAACCCACCAACCAACCAATCAACCAACCAAGAAATTTCCAAATGA 2874
Db 2898 AAAGCAGATCCCAAGAAATTAACAGCGCAGCCAAACCAAG--AAAATTCACAAAGAA 2956
QY 2875 GAAATGATTTGTCTTTTGCATTTTGTGTATTAAGCAATCATATTTACGAAATGAT 2934
Db 2957 GAATGGA---CTCTCTTCATCTTGTATTAACCAACCAATTAATTAAGTGAATTTCT 3012
QY 2935 CTTTTCTTAAAAAATGAGAGAAAGTGAATTTTCAAGAGTTTGGCCAG 2994
Db 3013 TCTTCTTTAAAAATTTAAAGTGAAGAAATGCAAAATTTATAGAGGTTTGGCCAG 3072
QY 2995 GGGTTAAATTCACAGATTTTAAAGGAAACACACAG--AGAAGCTACCTCAG 3051
Db 3073 GGGTTAAATTTACAGATTTTAAAGGAAACACACAAATTAAGGCTACCTCAG 3132
QY 3052 GTG-TTTTAACTCAGACCTTGTCTTGTGTTTCCCTTAAAGATTTGTAAGCTGATA 3110
Db 3133 GTGTTTAACTCAGACCTTGTCTTGTGTTTCCCTTAAAGATTTTGTAAACCTGATA 3192
QY 3111 GTTGAAGCATTTTATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3170
```

Db 3193 GTTGAGACA-TTTTATTTTAAATAATAGTT-GAAAAAATAATGAGC 3250

Qy 3171 TGCAGAGCTGGAGAGGTGACAGTCCAAGTGTGCAAGCTGTTCTGATTTGCTTCGCG 3230

Db 3251 TGCAGAGCTGGAGAGGTGACAGTCCAAGTGTGCAAGCTGTTCTGATTTGCTTCGCG 3305

Qy 3231 TAGCCAAAGACCAATATGAGCTTCTTTTGGCAAACTTGAATTTTATTT 3283

Db 3306 TAGCCAAAGACCTATGAGCTTCTTTGGAACAACTTGAATTTTATTT 3358

RESULT 2
AK077404 3262 bp mRNA linear HTC 18-SEP-2003
LOCUS Mus musculus 6 days neonate head cDNA, RIKEN full-length enriched library, clone:5430439D08 product:HEPATOCELLULAR CARCINOMA
DEFINITION AUTOANTIGEN homolog [Homo sapiens], full insert sequence.
AK077404 AK077404.1 GI:26097330
VERSION AK077404.1 GI:26097330
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

REFERENCE
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komo, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

REFERENCE
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komo, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M., Suni, N., Ishi, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujimake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, F., Watahiki, M., Yokeda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexed sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861

REFERENCE
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)

REFERENCE
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)

REFERENCE
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishi, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, U., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,

Qy 3171 TGCAGAGCTGGAGAGGTGACAGTCCAAGTGTGCAAGCTGTTCTGATTTGCTTCGCG 3230

Db 3251 TGCAGAGCTGGAGAGGTGACAGTCCAAGTGTGCAAGCTGTTCTGATTTGCTTCGCG 3305

Qy 3231 TAGCCAAAGACCAATATGAGCTTCTTTTGGCAAACTTGAATTTTATTT 3283

Db 3306 TAGCCAAAGACCTATGAGCTTCTTTGGAACAACTTGAATTTTATTT 3358

COMMENT
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: <http://genome.gsc.riken.go.jp/>
URL: <http://fantom.gsc.riken.go.jp/>
URL: <http://location/Qualifiers>

FEATURES
source 1..3262
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM,DB:5430439D08"
/db_xref="MG1:2393565"
/db_xref="taxon:10090"
/clone="5430439D08"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="6 days neonate"
misc_feature 16..1796
/note="HEPATOCELLULAR CARCINOMA AUTOANTIGEN homolog [Homo sapiens] (SPTR:Q9Y6M1, evidence: FASV, 88.3%id, 100%length, match=1775)
polyA_signal 3249..3254
/note="putative"
polyA_site 3262
/note="putative"

ORIGIN
Query Match 53.8%; Score 1765.4; DB 11; Length 3262;
Best Local Similarity 78.1%; Pred. No. 0;
Matches 2581; Conservative 0; Mismatches 466; Indels 256; Gaps 28;

Qy 58 GGAAGAGAGCGGATGATGAGCAAGCTTTCATCGGGAACCTGAGCCCGCTCACCGCC 117

Db 4 GGAAGAGCGGATGATGAGCAAGCTTTCATCGGGAACCTGAGCCCGCTCACCGCC 63

Qy 118 GACGACCTCCGCGACGCTTTTGGGACAGGAAGCTGCCCTGCGGACAGGTCTGCTG 177

Db 64 GACGACCTCCGCGACGCTTTTGGGACAGGAAGCTGCCCTGCGGACAGGTCTCTAC 123

Qy 178 AAGTCGCGCTACGCTTCTGTCGATCCCGGACGAACTGGGGCATCCGCGCATCGAG 237

Db 124 AAGTCGCGCTACGCTTCTGTCGATCCCGGACGAACTGGGGCATCCGCGCATCGAG 183

Qy 238 ACCCTCTGGGGTAAAGTGAATTCATGCGGAATTCATGGAATTCATGCACTCTCT 297

Db 184 ACCCTCTGGGGTAAAGTGAATTCATGCGGAATTCATGGAATTCATGCACTCTCT 243

Qy 298 AAAAGCTAAGAGCAGGAATTCATGATTCGAAATCCCTCTCTCACTGCGAGTGGAG 357

Db 244 AAAAGCTAAGAGCAGGAATTCATGATTCGAAATCCCTCTCTCACTGCGAGTGGAG 303

Qy 358 GTTGTGATGAGCTT-TTGGCTCAATATGAGGACATGGGAATGTGGAACAAGTCAAC 416

Db 304 GTTGTGATGAGGCTTGTGCTCAATATGAGGACATGGGAATGTGGAACAAGTCAAC 363

Qy 417 AGACACAGAAACCGCGTTCTCAACGTCATATGCAACAAGAGAAAGCAAAATATGC 476

Db 364 AGATACAGAAACCTCGCTTCTCAACGTCATATGCAACAAGAGAAAGCAAAATATGC 423

QY	477	CATGAGAAAGCTAAGCGGCGCATAGATTGGAACCTACTCCTTCAAGATTTCTCATCC	536
Db	424	TATTGAAAGCTCAGTGGGCGCATAGTTTGAGACTACTCCTTCAGATTTCTCATCC	483
QY	537	GGATTGAAGAGTGAAGCTCCCTTCGCCCCCTGACGAGCCGCGTGGGAGCACTCTTC	596
Db	484	CGATTGAAGAGTGAAGCTCTCTCACCCTCCATCG-----TGC	522
QY	597	CCGGAGCAAGGCGACGCGCTTGGGGCACTTCTGAGGCGAGACAGATTGATTTCCCGCT	656
Db	523	CCGGAAACAAGGCGACGCGCCCGGGAGCTTCTCGGCGCAGACAGATTGATTTCCCGCT	582
QY	657	GCGGATCTGGTCCCAACCCAGCTTTTGGTGCATCATCGAAAGAGAGGCGCTTGACAT	716
Db	583	GCGGATCTGGTCCCAACCCAGCTTTTGGTGCATCATCGAAAGAGAGGCGCTTGACAT	642
QY	717	AAAGAACATCACTAAGACAGCCAGTCCCGGGTAGATATCCATAGAAAAGAACTCTGG	776
Db	643	AAAGAACATCACTAAGACAGCCAGTCCCGGGTAGATATCCACAGAAAGAACTCTGG	702
QY	777	ACCTGAGAGAAAGCTGTCAACCATCCAGATCCAGCCGAGGGGAGCTTCTGAAGCATCGC	836
Db	703	GGCTGAGAGAAAGCTGTCAATCCATCTACATCCGAGAGGACATCTGAAGCATCGC	762
QY	837	CATGATCTTGAATCATGACAGAAAGGACAGATGAGACCAACTAGCCGAAGATTC	896
Db	763	CATGATCTTGAATCATGACAGAAAGGACAGATGAGACCAACTGGCTGAGAGATTC	822
QY	897	TCTGAAATCTTGGCACACATGGCTTGGTGGAGACTGATTGGAAAAGAGGACAAA	956
Db	823	TCTGAAATCTTGGGCCACATGGCTTGGTGGAGACTGATTGGCAAAGAGGACAAA	882
QY	957	TTTGAAGAAATTTGAACATGAAACGAGGACCAAGATACATCTATCTTGCAGAGATT	1016
Db	883	CCTGAAGAAATTTGAACATGACAGGAGCAAGATACATCTATCTTGCAGAGATT	942
QY	1017	GAGCATATACAACTCCGGAAGAAACCATCACTGTGAAGGCGACAGTTGAGGCTGTGG-CCA	1075
Db	943	GAGCATATTAACCCCGAGAGAACATCACTGTGAAGGCGACATTTGAAGCTGTGGCCA	1002
QY	1076	GTGCTGAATAGAGATTATGAAAGAGCTGCGTAGGCGCTTTGAAATGATATGCTGGCTG	1135
Db	1003	ATGCTGAATAGAGATTATGAAAGAGCTGCGTAGGCGCTTTGAGAAACAGACATGCTGGCG	1062
QY	1136	TTAAC-----	1140
Db	1063	TTAACCAACAGCCATCTGATCCAGGCGTAAACCTCAGTGCACCTTGGCATCTTTTGA	1122
QY	1141	-----	1140
Db	1123	CTGACCTGTGTGCTTCTCCACGAGAGGCCCCCGTAGGATTCCTCCCACTCTCCT	1182
QY	1141	-----ACCACTCCGAGTACTTCTCCAGCTGTACCCCATCAGCATTTGG	1186
Db	1183	ATCACCCCTTTGCTAACCCACTCCGAGTACTTCTCCAGCTGTACCCCATCAGCATTTGG	1242
QY	1187	GCGCGTTCCGGATATCACTCTTAACAGAGCGAGAGATTTGATTTCTTCACTCCCA	1246
Db	1243	GCGCATTCACATATATCACTCTTAACAGAGCGAGAGATTTGATTTCTTCACTCCCA	1302
QY	1247	CCAGAGCTGTGGCGCCATCATCGGGAAGAAAGGGGCGACATCAATCAACAGCTGCGAGAT	1306
Db	1303	CCAGAGCTGTGGGCTCATTCATCGGGAAGAAAGGGGCGACATCAATCAACAGCTGCGAT	1362
QY	1307	TGCGCGAGCTCTATCAAGATTGCGCTGCGGAAGGCCAGACGTCAAGCAAGATGG	1366
Db	1363	TTGCTGGTGCCTTCATCAAGATTGCTCAGAGAGAGTCCAGATGTCAGTGAGAGATTGG	1422
QY	1367	TCATCATACGCGGCGAACCGGAAGCCCATGTCAGAGCCAGGATCTTTGGGAAAC	1426
Db	1423	TCATCATACGCTCTCTCTAAGCCCATTTAAGGCTCAGAGCGGATCTTTGGGAAAC	1482

QY	1427	TGAAAGAGAAAACTTCTTTAACCCCAAGAAGACTGAAGCTGGAGCGCATATACAG	1487
Db	1483	TGAAGAGAAAACTTCTTTAATCCCAAGAAAGTAGAAGCTGGAGGCCCATCTCGAG	1544
QY	1487	TGCCCCCTTCACAGGTGGCCGGTGATTTGGCAAAAGTBGGCAAGACCGTGAACGAATTGC	1544
Db	1543	TCCCCTGTCGACCGGTGGCCGGTGATTTGGCAAGGCGGCAAAAACCGTGAACGAGCTGC	1602
QY	1547	AGAACTTAAACAGATGACAAAGTCATGTGCTCCTGTAACCAAACCGCCAGATGAAATGAAG	1602
Db	1603	AGAACTTGAACAGTGAAGAGATTTATGTGCTCTGTACCAAACCGCACAGAGAAATGAAG	1667
QY	1607	AAGTATTCCTCAGAAATTATCGGACCTTCTTTGTAAGCCAGACTGCACGCGCAAGATCA	1667
Db	1663	AAGTATTCCTCAGAAATTATCGGACCTTTTTTGTATGACGAGCTGCACAACCGCAAGATCA	1722
QY	1667	GGAATAATTGTAACAACAGGTGAAGGACAGAGAGACGAATAACCTTACGGAGTGTGCTCAC	1722
Db	1723	GGAATAATTGTAACAGAGGTGAAGGACAGAGAGACGAAGTACCCTCAGGGAGTGTGCCCCAC	1787
QY	1727	AGGCGACGAAGTAGGCTCCACAGGACACACAAACAAAGGATGAATGTAGCCCTTCC	1787
Db	1783	AGGCGACGAAGTAGGCTCCACAG--CACACAGACGAACCGATGAATGTAGCCCTTCC	1844
QY	1787	AACACTGACAGATGAGCCAAACGACGCCAGATCGGAGCAAAACCAAGACAT	1844
Db	1841	AACACTGACAG-ATGAGACCAAC--AGCCAGCAGATCGGAGCAAAACCAAGACAT	1897
QY	1847	-CTAGGAAATGAAAGTCTGCGGAGAGCGGCCACGGACCTGCGGAGGCCCTGAGAACCC	1907
Db	1897	CCCGAGAGATGTGCAAGTCTGCAGACACAGCCAGGACCT---GCAGACCTTACACATCC	1952
QY	1906	AGGGGCCGAGAGGGGCGGGGAAAGGTCAAGCAGGTTTGCAGAACCAACCG--AGCCCGG	1966
Db	1953	TGGGATTCAGAGGGGACAGGGAA--GCCAGTTTGTCCAGAACACCGCTTGGCTTGC	2005
QY	1964	CTCCGCGCCCCAGGGCTTCTGCAGGCTT--CAGCCATCCAATTACCATCCACTCGGAT	2022
Db	2010	CCCCAGCTTCCCTGGCTTCTGCAGGCAATTAACGCCATTCACCTGCCATCCACATCBAAT	2066
QY	2022	CTCTCTCTGAATCCCAAGACGCTATCCCTTTAGTTGAATPAACATAGGTGA--CGTGT	2079
Db	2070	-TCTCTCAGTCCCAAGACGCTATCCCTTTCGTTGAACPAACATAGGTGAACATGCT	2128
QY	2080	TCAAAGCCAAAGAAATGACACCCCTTTTC-----TGTGCAAAATGCTCTGTGACATG	2134
Db	2129	CAAAGCCAAAGAAAAATTCCTAGCCGTTTTCTTTGTGTGAAGATGTCTCTGTACATG	2188
QY	2135	TGTGTACATATTAGAAAGGAA--GATGTTAAGATATGTGCTGTGGTTACACAGGCTG	2193
Db	2189	TATGTACATATCAAGAGGGGAAGATGTTAAGAAATGTGCTGTGGTTACACAGGCTG	2248
QY	2194	CCTGACGGGTATATATTTTGAAGATATATCAAAATCTCAACTACTCCAAATTT	2253
Db	2249	CCTGACGGGTATATATTTTGAAGAAATATATCAAAATCTCAACTCAAAATTTT	2308
QY	2254	TAAATCAATATTAATTTTTTTCTTTTAAAGGAAAGAGGCTTTTCTGACTTTAA	2313
Db	2309	TACTCAGTATTAATTTTTTTCTTTTGAAGGAAAGAGGCTTTTCTGACTTTCAA	2368
QY	2314	GAATTAAGCTCTTGGAGGCTCTCAAGCT-----GTAGAGAGAGCTTTGAAGGCCAC	2365
Db	2369	GAATTAAGTATTTGGGTGAGGCTCCCACTTGGTATTAAGAAAGGAACTTTGAAGGCCAC	2428
QY	2366	CGCAGAAA-----TTCAACCAAGAGGAAATCTGTGCGAA	2401
Db	2429	CACACAAAAAATTTCACTTGGAAAGGAAAAATTTTTTCCAAAGCACTTTAATTGCC	2488
QY	2402	GGAACCTACAGGCAAGTTGTGGAACACCGTGTATGTAACAAGAAAGGATACCGTCTCTT	2461
Db	2489	AGTTTGTGGTTCAACCTGTCTTTTCTTTGTCCAGACGAAAAAGGAAAAACCTTCTCTCT	2548
QY	2462	GAAAGGAACTCTGTCACTCTCATCCTGTCTAGCTCAATACCACTTCTCTTGTCT	2521

Db	2549	TAAAGGAAAGGAGCCTTAGAGCCCTTGTCACCCCTGTCTTCCACACACCACATGCTCAGTT	2608
QY	2522	TCACAGGTTTAAAC-----TGTTTTTGCATCTGCT	2555
Db	2609	TCACAGGTTTTTAAATTAAATTATTTTGGTTGGTTTGGTTTGGTTTGGTATCTAGCT	2668
QY	2556	AATTAATCTCTGTCTCTCTCTGTGTATCTCTCCCTCCCTCCCTCCCTCTCTCCCA	2615
Db	2669	AATATAT--TCTGTCTCTCTCTTATCTCTTAACTTCTCTCTCTCTTAACTCC	2726
QY	2616	TCATCCATCTCTTGAATTTCC--TCATCCCTCATCTCATCCCGATCTAGCACC--	2670
Db	2727	CAGAAATTTCTTTTGTAGTCCCTTGTGTTCTTTTCTGTCTCCCGATATAGACCCC	2786
QY	2671	-----CCCCCCCCCAGGCAAGACAGTGTCTGTAGTATCATCATACAAAGAAACAAAGC	2729
Db	2787	TCCCTCTCCCCCGAGGCAAGACAGTGTCTGAATATCATACACACAAAGAAACAAATGC	2846
QY	2730	GAAACACACAAACACAGCTCTACTTACACTTGTGTACTGAAAGAAAGAAAGTCAATGT	2789
Db	2847	GAAACACACAGACACAGCCTTAGCTTACACTGTGTACTGAAAGAAAGAAAGTGTGCGGT	2906
QY	2790	ACTGTGCTTAGCGTTTGTGAAGAGAAACAGGAAACCAACCAACCATCAACCA	2849
Db	2907	ACTGTGCTTAGCG--ATTGGAAGAGAAAGAAAGATCCACCAATTAACACGACGACCA	2965
QY	2850	ACAAAGAAAAATTCACATGAAAGAAAGATATTTGTCTTTTGCATTTTGTGTATTA	2909
Db	2966	ACAAAG--AAAAATCCCAAGAAAGAAAGATGA-----CTCTCTCTCCATTTGTGTATCAA	3020
QY	2910	GCCATCATATTTACGCAAAATGATTTCTTTCTTTAAAAAAAATGTGAGAGAAATAGC	2969
Db	3021	GCCATCATCATTTATGTGAAATTTCTTTCTTTAAAAAAAATTTAAAGTCAGAGAAATAGC	3080
QY	2970	AAATTTTCCAGAGTGTGTGGCCGAGGCGTTAAATTCACAGATTTTAAAAAGAA	3029
Db	3081	AATTTATACAGGTTGTGGCCGAGGCGTTAAATTTACAGATTTTAAAAAGAGAA	3140
QY	3030	CACACAGA--AGAAAGTACCTCAGGTG--TTTTACCTCAGACCTTGCTCTGTGTTTC	3085
Db	3141	AACACAAATTAAGAGTACCTCAGGTGTTTAACTCAGACCTTGCTCTGTGTTTC	3200
QY	3086	CCTTAGAGATTTGTAAAGCTGATATGTTGACGATTTTATTTTAAATAAATGA	3145
Db	3201	CCTTAGAGATTTGTAAAGCTGATATGTTGAGCA--TTTTTATTTTAAATAAATGA	3259
QY	3146	GTT 3148	
Db	3260	GTT 3262	
RESULT 3			
LOCUS AK044984			
DEFINITION Mus musculus 9.5 days embryo parthenogenote cDNA, R1EN full-length			
AUTOMATIC ENUCLIC, clone:Bi30017M24 Product:HEPATOCELLULAR CARCINOMA			
ACCESSION AK044984			
VERSION AK044984.1 GI:26090647			
KEYWORDS HTG; CAP trapper.			
SOURCE Mus musculus (house mouse)			
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE 1 Carninci, P. and Hayashizaki, Y.			
TITLE High-efficiency full-length cDNA cloning			
JOURNAL Meth. Enzymol. 303, 19-44 (1999)			
MEDLINE 99279253			
PUBMED 10349636			
REFERENCE 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,			
AUTHORS			

TITLE	Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
PUBMED	20493374
REFERENCE	11042159
AUTHORS	3 Shibatani, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Komno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Isikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE	RIKEN Integrated sequence analysis (RISA) system--384-format
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	4
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
REFERENCE	5
AUTHORS	The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
REFERENCE	6 (bases 1 to 29664)
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagasaki, T., Hara, A., Hashizume, M., Hayashida, S., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondou, S., Komno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, R., Murata, M., Nakamura, Y., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, S., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Shuho-cho, Tsukuba-ku, Yokohama Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in RIKEN Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
	Tissues were provided by Dr. Tomohiro Kono (Department of Animal Science, Tokyo University of Agriculture, 1737 Hnateko Atsugi City, Kanagawa Prefecture, Japan) whose assistance we gratefully acknowledge.
FEATURES	Please visit our web site for further details.
SOURCE	URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/ Location/Qualifiers 1. 2964 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="PANTOM:DB:B130017M24" /db_xref="MGI:2410341" /db_xref="taxon:10090" /clone="B130017M24" /issue_type="parthenogenote"

misc_feature
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="9.5 days embryo"
1..2964
/note="HEPATOCELLULAR CARCINOMA AUTOANTIGEN homolog [Homo
sapiens] (SPTB|Q9Y6M1, evidence: PASTY, 88.3%ID,
100%length, match=1775)"

ORIGIN

Query Match 49.4%; Score 1622; DB 11; Length 2964;
Best Local Similarity 78.2%; Pred. No. 4.2e-307;
Matches 232; Conservative 0; Mismatches 435; Indels 216; Gaps 23;

```
QY 247 GGTAAAGTGAATTCATGCGAATAATCATGAAAGTTTACTGATCTCTTAAAGCTA 306
Db 30 GGTAAAGTGAATTCATGCGAATAATCATGAAAGTTTACTGATCTCTTAAAGCTA 89
QY 307 AGGAGCGAAGAAATTCATGCTGAAACATCTCTCTCTCACTGCGATGGAGGTGGAT 366
Db 90 AGGAGCGAAGAAATTCATGCTGAAACATCTCTCTCTCACTGCGATGGAGGTGGAT 149
QY 367 GGAATTTGGCTCAATATGGAGAGTGGAGAAATGGAAAGTCAACAGACAGACGAA 426
Db 150 GGAATTTGGCTCAATATGGAGAGTGGAGAAATGGAAAGTCAACAGATACGAA 209
QY 427 ACCGCGCTTGCAACGTCAATATGCAACAAGAGAGCAAAATAGCCATGAGAG 486
Db 210 ACTGCCCTTGCAACGTCACTATATGACAGAGAGAGGCAAGCTATAGATGAGAG 269
QY 487 CTAAAGCGGAGATCATGTTGGAGAACTACTCTTCAAGTTTCTCAATCCGAGTAAAG 546
Db 270 CTAAAGCGGAGATCATGTTGGAGAACTACTCTTCAAGTTTCTCAATCCGAGTAAAG 329
QY 547 GTGAGCTCCCTTGCCCGCTCAGCGAGCCAGGCTGGGAGCACTTTCCCGGAGCA 606
Db 330 GTGAGCTCCCTTGCCCGCTCAGCG-----TGCCCGGAGCA 368
QY 607 GGCACGCGCCCTGGGGGCACTTCTCAGCGCAGACAGATTATTTCCGCTCGGATCTG 666
Db 369 GGCACGCGCCCTGGGGGCACTTCTCAGCGCAGACAGATTATTTCCGCTCGGATCTG 428
QY 667 GTCCCAACCAATTTGTTGGTGCATATGAGAAAGAGGGCTTGAACATTAAGAACATC 726
Db 429 GTCCCAACCAATTTGTTGGTGCATATGAGAAAGAGGGCTTGAACATTAAGAACATC 488
QY 727 ACTAAGACAGCCAGTCCGGGTAGATATCCATTAAGAAAGAACTCTGAGCTGACAG 786
Db 489 ACTAAGACAGCCAGTCCGGGTAGATATCCATTAAGAAAGAACTCTGAGCTGACAG 548
QY 787 AAGCTGTCAACATCCATGCAACCCAGAGGGGACTTCTGAACATGCGCATGATCTT 846
Db 549 AAGCTGTCAACATCCATGCTACCCAGAGGGGACATCTGAACATGCGCATGATCTT 608
QY 847 GAAATCATGCAAGAAAGGCGAGATGAGACCAATCTAGCCGAGAGATTCTCTGAATTC 906
Db 609 GAAATCATGCAAGAAAGGCGAGATGAGACCAATCTAGCCGAGAGATTCTCTGAATTC 668
QY 907 TTGGCAACAATAGCTTGGTTGGAAGACTGATTGAAAGAAAGAGCAATTTTGAAGAA 966
Db 669 TTGGCAACAATAGCTTGGTTGGAAGACTGATTGGAAGAAAGAGCAATTTTGAAGAA 728
QY 967 ATTGAACATGAAACAGAGGACCAAGATTAACATCTCATCTTTGAGAGATTGACATATAC 1026
Db 729 ATTGAACATGAAACAGAGGACCAAGATTAACATCTCATCTTTGAGAGATTGACATATAC 788
QY 1027 AACCCGGAAGAAACATCTGCTGAAAGGCGACGTTGAGGCTGTGCCAGTGTGAGATA 1086
Db 789 AACCCGGAAGAAACATCTGCTGAAAGGCGACGTTGAGGCTGTGCCAGTGTGAGATA 848
QY 1087 GAGATTATGAAGAGCTCGAGAGGCTTTGAGAAATGATATGCTGCTGTAAAC----- 1140
Db 849 GAGATTATGAAGAGCTCGAGAGGCTTTGAGAAATGATATGCTGCTGTAAACAA 908
QY 1141 ----- 1140
```

```
Db 909 GCCAATCTGATCCCAAGGCTTAACCTCAGTGCACCTTGGCATCTTTGACTGACTGTCT 968
QY 1141 ----- 1140
Db 969 GTGCTTCTCCACAGCAGAGGCCCCGTGAGATTCGCCCACTCTCTATCACCCCTTT 1028
QY 1141 ----ACCCATCTCCGATATCTTCTCCAGCTGTACCCCATCACAGTTTGGCCGCTTCCG 1197
Db 1029 GCTAACCACTCCGATATCTTCTCCAGCTGTACCCCATCACAGTTTGGCCCATTCACA 1088
QY 1198 CATCATCACTCTTACAGAGCAGAGATGTGATCTCTTCAATCCCAACCCAGCTGTG 1257
Db 1089 CATCATCACTCTTACAGAGCAGAGATGTGATCTCTTCAATCCCAACCCAGCTGTG 1148
QY 1258 GCGCCATCATCTGGGAAGAAAGGGGCAACATCAAAAGCTGGCGAGATTTGCCCGAGCC 1317
Db 1149 GGTGCTATCATCTGGGAAGAAAGGGGCAACATCAAAAGCTGGCGATTTGCTGTGTCC 1208
QY 1318 TCTATCAAGATGCTCCCTGCGGAAGGCCCAAGCTGACGAGAGATGTGATCATCAC 1377
Db 1209 TCTATCAAGATGCTCCCTGCGGAAGGCCCAAGCTGACGAGATGTGATCATCATCT 1268
QY 1378 GGGCCACCGGAAGCCAGTTCAAGGCCCAGAGGACGATCTTTGGGAAATGAAAGAGAA 1437
Db 1269 GGTCTCTCTGAAGCCAGTTTAAAGCTCAAGAGCGATCTTTGGGAAATGAAAGAGAA 1328
QY 1438 AACTCTTTTAAACCCCAAGAGAAAGTGAAGCTGGAACGATATCAAGTGCCTCTTCC 1497
Db 1329 AACTCTTTTATCCCAAGAGAAAGTGAAGCTGGAAGCCCAATCCAGATGCCCTCTG 1388
QY 1498 ACGCTGGCCGGGTGATTGGCAAAAGTGGCAAGACGTGAACGAATCTGCAAACTTAAC 1557
Db 1389 ACGCTGGCCGGGTGATTGGCAAAAGTGGCAAAAGTGGCAAAAGTGGCAAACTTGA 1448
QY 1558 AGTGAAGAAATCATCTGCTCGTGAACCAACCGCATGAATGAAGAGATGCTC 1617
Db 1449 AGTGAAGAAATCATCTGCTCGTGAACCAACCGCATGAATGAAGAGATGCTC 1508
QY 1618 AGAATTATCGGGCACTTCTTTGCTAAGCTAGACTGCAAGGCAAGATCAAGGAAATTGTA 1677
Db 1509 AGAATTATCGGGCACTTCTTTGCTAAGCTAGACTGCAAGGCAAGATCAAGGAAATTGTA 1568
QY 1678 CAACAGGTGAAGCAGAGAGGCAAAATCTTGAAGAGTCCCTCAACGCGCAGAG 1737
Db 1569 CAACAGGTGAAGCAGAGAGGCAAAATCTTGAAGAGTCCCTCAACGCGCAGAG 1628
QY 1738 TGAGGCTCCACAGGCAACCAACGATGAATGAAGCTTCCCAACCTTGACA 1797
Db 1629 TGAGGCTCCACAGG--CACTAGCAAGCAACCGATGAATGAAGCTTCCCAACCTTGACA 1686
QY 1798 GAATGACCAACCAACGAGCCAGCCAGATCGGAGCAACCAAGACCAT--CTGAGAGATG 1856
Db 1687 G-ATGAGACCAAC--AGCAGCAATATGAGGAGCAACCAAGAGCATCCAGAGAGTG 1742
QY 1857 AGAATCTGCGAGGCGCGCGAGGAACTTGCGAGGCTTGAAGACCCAGAGGCGAGG 1916
Db 1743 CGAGATCTGAGAGCAAGCGAGGCGCTGC-----AGACTCTTCAACATCTGGGATCAGG 1797
QY 1917 AGGAGCGGGAAGGTGAGCGAGGTTGCGAGAACCAACG--AGCCCGGCTCCGCGCCCG 1974
Db 1798 AGGAGCAAGGAG--GCCAGGTTTCCAGAAACCGGTTGGCTGCGCCCAAGCTTCC 1854
QY 1975 CAGGGCTTGAAGGCTT--CAGCCATCACTTCAACATCACTCGGATCTCTCTGAAC 2032
Db 1855 CCGGCTTCTGAGAGGATTAACAGCCATCACTGCGCATCACTCATGAT--TCTCTCAGT 1913
QY 2033 TCCCAAGAGCTATCTCTTTTGTGTAATTAATATGTAACGTGTTCAAAGCAGCA 2092
Db 1914 TCCCAAGAGCTATCTCTTTTGTGTAATTAATATGTAACGTGTTCAAAGCAGCA 1973
QY 2093 AATGACACGCC-----TTTTGTGGCAAAATCGTCTGTATCATGTGTGATTTG 2148
```

Db 1974 AATTCCTGACCGCTTCTTCTGTGGAAGATGTCCTGACATGATGACATATCAG 2033
 QY 2149 AAGGGA--GATGTTAAGATATGCGCTGTGGTTACACAGGGTCCCTGACGGGTAT 2207
 Db 2034 AAGGGAAGATGTTAAGAAATGCGCTGTGGTTACACAGGGTCCCTGACGGGTAT 2093
 QY 2208 ATATTTTGAAGAT 2267
 Db 2094 ATATTTTGAAGAT 2153
 QY 2268 TTTTCTTTCTTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2327
 Db 2154 TTTTCTTTCTTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2213
 QY 2328 G-GAGGTCACGGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2385
 Db 2214 GTGAGTCACGGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2273
 QY 2386 GGAATCTCGTCGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2445
 Db 2274 GGAATCTCTTCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2332
 QY 2446 GGAATCTCTTCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2505
 Db 2333 GGAATCTCTTCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2392
 QY 2506 CCCATTTCTCTT-----TGTTCACAGGTTTAACTGTTT 2545
 Db 2393 TTGTCAGGTTTCAAGGTTTAACTGATGATGATGATGATGATGATGATGATGAT 2452
 QY 2546 GCATCTGCTAT 2605
 Db 2453 GCATCTGCTAT 2506
 QY 2606 TTCTTCTCATCTCATCTTCTTGAATTTCTCATCTCATCTCATCTCATCTCATCT 2665
 Db 2507 GTCTTCTTCAACCAAAATCTTTTGAATCTTCTTCTTCTTCTTCTTCTTCTTCTT 2566
 QY 2666 GATC-----CCCCCCCCCAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2722
 Db 2567 GATC-----CCCCCCCCCAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2626
 QY 2723 CAAAAGCAG 2782
 Db 2627 CAAATGCGAAG 2686
 QY 2783 CAATGCTATCTTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2842
 Db 2687 CGGCGGTATCTTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2745
 QY 2843 CAACCAAG 2902
 Db 2746 CAGCCCAAG 2800
 QY 2903 TGTATAGGCTAT 2962
 Db 2801 TATACAGGCTAT 2860
 QY 2963 AAAGTGAAGATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3022
 Db 2861 AAATAGCAATTTTATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2920
 QY 3023 AGAAAAAG 3059
 Db 2921 TGAGAAAAAG 2960

RESULT 4
 BM561242 999 bp mRNA linear EST 20-FBB-2002
 LOCUS BM561242 6567353 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5739550
 DEFINITION AGENCOURT 6567353 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5739550
 5', mRNA sequence.
 ACCESSION BM561242

VERSION BM561242.1 GI:18806337
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 999)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at:
 http://image.llnl.gov
 Plate: LHAM12753 row: a column: 23
 High quality sequence stop: 679.
 FEATURES
 source location/Qualifiers
 1. 999
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5739550"
 /tissue_type="duodenal adenocarcinoma, cell line"
 /lab_host="VDH108 (phage-resistant)"
 /clone_lib="NIH-MGC 88"
 /note="Organ: small intestine; Vector: pCMV-SPORT6; Site 1: Not1; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.767 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH-MGC Library."
 ORIGIN
 Query Match 28.7%; Score 941.6; DB 12; Length 999;
 Best Local Similarity 98.1%; Pred. No. 7.9e-174;
 Matches 983; Conservative 0; Mismatches 16; Indels 3; Gaps 3;
 QY 1029 CCGGAAAGAACATCTGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1088
 Db 1 CCGGAAAGAACATCTGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
 QY 1089 GATTATGAAG 1148
 Db 61 GATTATGAAG 120
 QY 1149 CGGATCTTCTCAGCCTGTACCCCATCAGAGAGAGAGAGAGAGAGAGAGAGAG 1208
 Db 121 CGGATCTTCTCAGCCTGTACCCCATCAGAGAGAGAGAGAGAGAGAGAGAGAG 180
 QY 1209 TTATCAG 1268
 Db 181 TTATCAG 240
 QY 1269 CGGAAAG 1328
 Db 241 CGGAAAG 300
 QY 1329 TGCCTCTGCGAAG 1388
 Db 301 TGCCTCTGCGAAG 360
 QY 1389 AGCCAGGTTCAAG 1448
 Db 361 AGCCAGGTTCAAG 420
 QY 1449 CCCCAAG 1508
 Db 421 CCCCAAG 480

```
QY 1509 GGTGATTGGCAAAAGGTGGCAAGACCGTGAACGAACTGCAAGAACTTAACTGACAGT 1568
Db 481 GGTGATTGGCAAAAGGTGGCAAGACCGTGAACGAACTTAACTGACAGT 540
QY 1569 CATGTGCTCTGTGACCAAAAGCCGATGAAATAGAGAAATGATGCTGCAATATTATCG 1628
Db 541 CATGTGCTCTGTGACCAAAAGCCGATGAAATAGAGAAATGATGCTGCAATATTATCG 600
QY 1629 GCACTTCTTGTGCTGACGACCTGCAAGCGAAGATGAGGAAATTTGACAAACAGTGA 1688
Db 601 GCACTTCTTGTGCTGACGACCTGCAAGCGAAGATGAGGAAATTTGACAAACAGTGA 660
QY 1689 GCAAGCAGAGAGCAAAATACCTCTCAAGAGTGCCTCAAGCGCAAGCAAGTGAAGCTCCA 1748
Db 661 GCAAGCAGAGAGCAAAATACCTCTCAAGAGTGCCTCAAGCGCAAGCAAGTGAAGCTCCA 720
QY 1749 CAGGCAACCAAGCAAAACCAAGGATGATGATGAGCTTCCACACCTGACAGATGACCA 1808
Db 721 CAGGCAACCAAGCAAAACCAAGGATGATGATGAGCTTCCACACCTGACAGATGACCA 780
QY 1809 AACGACACCAAGCAGATCGGAGGCAAAACCAAGACCACTTGAGGAAATGAGAACTTCCG 1868
Db 781 AACGACACCAAGCAGATCGGAGGCAAAACCAAGACCACTTGAGGAAATGAGAACTTCCG 839
QY 1869 AGGCGGCAAGGACTCTGCGGAGCCCTGAGAAACCCCAAGGAGCCGAGAGGAGGAGGAA 1928
Db 840 AGGCGGCAAGGACTCTGCGGAGCCCTGAGAAACCCCAAGGAGGAGGAGGAGGAGGAA 898
QY 1929 GGTACAGCAGGTTTGGCAAGACCAAGGAGCCGCTCCGCGCCGCCCAAGGAGCTTCCAG 1988
Db 899 AGTACACCCAGGTTTGGCAAGACCAAGGAGCCGCTCCGCGCCGCCCAAGGAGCTTCCAG 958
QY 1989 GCTTACAGCAGTCACTTCCACATCAGTCCGAGTCTCCCTGA 2030
Db 959 GCTTACAGCAGTCACTTCCACATCAGTCCGAGTCTCCCTGA 999
```

```
RESULT 5
LOCUS B0691878
DEFINITION AGENCOURT_8034698 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6207026
5', mRNA sequence.
ACCESSION B0691878
VERSION B0691878.1 GI:21817194
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM2362 row: h column: 03
High quality sequence stop: 653.
Location/Qualifiers
1. 874
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6207026"
/risue_type="dicta carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_110"
```

FEATURES

```
source
```

/note="Organ: pancreas; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."

Query Match 25.7%; Score 842.2; DB 13; Length 874; Best Local Similarity 99.4%; Pred. No. 2.3e-154; Matches 856; Conservative 0; Mismatches 3; Indels 2; Gaps 1;

```
QY 1695 GAGCAGAAATACCTCTGAGGAGTGCCTCAAGCGACAGCAAGTGAAGGCTCCACAGGCA 1754
Db 1 GAGCAGAAATACCTCTGAGGAGTGCCTCAAGCGACAGCAAGTGAAGGCTCCACAGGCA 60
QY 1755 CCGCAAAACCAAGGATGATGATGAGCTTCCACACCTGACAGATGACCAAGCA 1814
Db 61 CCGCAAAACCAAGGATGATGATGAGCTTCCACACCTGACAGATGACCAAGCA 120
QY 1815 GCCAGCCAGATCGGAGCAAAACCAAGACCACTTGAGAAATGAGAACTTCCGAGGCGG 1874
Db 121 GCCAGCCAGATCGGAGCAAAACCAAGACCACTTGAGAAATGAGAACTTCCGAGGCGG 180
QY 1875 CCAAGGACTTCTCCGAGGCTTGAAGAACCCCAAGGAGCGAGAGGAGGAGGAGTCA 1934
Db 181 CCAAGGACTTCTCCGAGGCTTGAAGAACCCCAAGGAGCGAGAGGAGGAGGAGTCA 240
QY 1935 CCAAGGTTGCAAGACCAAGGAGGCTTCCGCGCCGCCCAAGGAGCTTCCAGGCTTCA 1994
Db 241 CCAAGGTTGCAAGACCAAGGAGGCTTCCGCGCCGCCCAAGGAGCTTCCAGGCTTCA 300
QY 1995 GCCATCCACTTCAACATCCACTGAGATCTCTCTGAACTCCCAAGAGCTATCCCTTTA 2054
Db 301 GCCATCCACTTCAACATCCACTGAGATCTCTCTGAACTCCCAAGAGCTATCCCTTTA 360
QY 2055 GTTGAACCTAACATAGGTGAAGCTGTTCAAAGCCCAAGCAAAATGACACCTTTTCTGTG 2114
Db 361 GTTGAACCTAACATAGGTGAAGCTGTTCAAAGCCCAAGCAAAATGACACCTTTTCTGTG 420
QY 2115 GCAAAATGCTCTGTACATGTGTATATTTGAAGGGAAGATGTTAAGATGTGCG 2174
Db 421 GCAAAATGCTCTGTACATGTGTATATTTGAAGGGAAGATGTTAAGATGTGCG 480
QY 2175 CTGTGGATTACAGAGGAGCTGCGAGGATATATTTTGAAGATATATATCAATTA 2234
Db 481 CTGTGGATTACAGAGGAGCTGCGAGGATATATTTTGAAGATATATATCAATTA 540
QY 2235 CTCACCTAACCTCAATTTTAAATCAATTTAATTTTCTTTTAAAGAAAGCA 2294
Db 541 CTCACCTAACCTCAATTTTAAATCAATTTAATTTTCTTTTAAAGAAAGCA 600
QY 2295 GGTCTTTCTAGACTTTAAAGATTAAGTCTTTGGAGGTCTCAGCGGTGAGAGAGCT 2354
Db 601 GGTCTTTCTAGACTTTAAAGATTAAGTCTTTGGAGGTCTCAGCGGTGAGAGAGCT 660
QY 2355 TTGAGGCAACCCGCAAAATCAACCAAGGGAATCTGTTGGAAGGACATCCAGCG 2414
Db 661 TTGAGGCAACCCGCAAAATCAACCAAGGGAATCTGTTGGAAGGACATCCAGCG 720
QY 2415 AGTTCTGATCACTGTGTATGCAACAGAGGATACCGTCTCTTGAAGAGAACTC 2474
Db 721 AGTTCTGATCACTGTGTATGCAACAGAGGATACCGTCTCTTGAAGAGAACTC 780
QY 2475 TGTCACTCTCATGCTGTGTAGCTATACACCATTTCTTTGCTTCAAGGTTTAA 2534
Db 781 TGTCACTCTCATGCTGTGTAGCTATACACCATTTCTTTGCTTCAAGGTTTAA 840
QY 2535 A--CTGATTTTGTGACTAG 2553
Db 841 AACTGGTTTTTGTGACTAG 861
```

RESULT 6
 BQ221568
 LOCUS
 DEFINITION
 AGNCOURT 7559207 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6046477
 5', mRNA sequence.
 BQ221568
 BQ221568.1 GI:20402968
 EST.
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 851)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgsbbs@mail.nih.gov
 Tissue Procurement: ATCC/DCTD/DTF
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LIML at:
 http://image.liml.gov
 Plate: LIML1292 row: f column: 14
 High quality sequence stop: 637.
 Location/Qualifiers
 1..851
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6046477"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_1db="NIH MGC 72"
 /note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2 kb. Library constructed by Life
 Technologies."

ORIGIN
 Query Match 25.3%; Score 830; DB 13; Length 851;
 Best Local Similarity 99.3%; Pred. No. 5.7e-152;
 Matches 844; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

47 GGGGCTCTCGGGGAAGAGACGATGATGACAGCTTTACATCGGGAAGCTGAGCCCG 106
 |||||
 1 GGGGGCTCTCGGGGAAGAGACGATGATGACAGCTTTACATCGGGAAGCTGAGCCCG 60
 |||||
 107 CCGTCAACCGCGACGACCTCCGACGCTTTGGGGACGGAAGCTGCCCTGGCGGAC 166
 |||||
 61 CCGTCAACCGCGACGACCTCCGACGCTTTGGGGACGGAAGCTGCCCTGGCGGAC 120
 |||||
 167 AGTCTCTGTAAGTCCGCTACGCTTGTGACATACCCGACGAACTGGGCATCC 226
 |||||
 121 AGGTCTGTGAAGTCCGCTACGCTTGTGACATACCCGACGAACTGGGCATCC 180
 |||||
 227 GGGCCATCGAGACCTCTCGGGGTAAGTGAATTCATGGGAAATCATGGAAGTTGATT 286
 |||||
 181 GGGCCATCGAGACCTCTCGGGGTAAGTGAATTCATGGGAAATCATGGAAGTTGATT 240
 |||||
 287 ACTGAGTCTTAAAAAGCTPAAGAGAGGAAATTCAGATTGCAACATTCCTCCCTACCC 346
 |||||
 241 ACTGAGTCTTAAAAAGCTPAAGAGAGGAAATTCAGATTGCAACATTCCTCCCTACCC 300
 |||||
 347 TGGAGGGAAGGTGTGATGACATTTGGCTCAATATGGAAGAGTGAATGTGAAC 406
 |||||
 301 TGGAGGGAAGGTGTGATGACATTTGGCTCAATATGGAAGAGTGAATGTGAAC 360
 |||||
 407 AAGTCAACAGACAGAAACCGCGTTGTCAACGTCAATATGACAAAGGAAG 466

Db
 361 AAGTCAACAGACAGAAACCGCGTTGTCAACGTCAATATGACAAAGGAAG 420
 |||||
 QY
 467 CAAAATAGCATGAGAGAGTAAAGCGGACATCAGTTTGAGAACTACTCTTCAAGATT 526
 |||||
 Db
 421 CAAAATAGCATGAGAGAGTAAAGCGGACATCAGTTTGAGAACTACTCTTCAAGATT 480
 |||||
 QY
 527 CTTACATCCCGGATGAAGAGGTGAGCTCCCTTGTGCCCCCTCAGCGAGCCAGGCTGGG 586
 |||||
 Db
 481 CTTACATCCCGGATGAAGAGGTGAGCTCCCTTGTGCCCCCTCAGCGAGCCAGGCTGGG 540
 |||||
 QY
 587 ACCACTCTTCCCGGAGCAAGGCGCAGCGCCCTGGGGGCACTTCAGGCGCAGAGATTG 646
 |||||
 Db
 541 ACCACTCTTCCCGGAGCAAGGCGCAGCGCTTGGGGGCACTTCAGGCGCAGAGATTG 600
 |||||
 QY
 647 ATTTCGCGCTGCGATCCTGTGTCGCCACCCAGTTTGTGTGTCATATCGAAAGAGG 706
 |||||
 Db
 601 ATTTCGCGCTGCGATCCTGTGTCGCCACCCAGTTTGTGTGTCATATCGAAAGAGG 660
 |||||
 QY
 707 GCTTACCATTAAGAAACATCACTAAGACACCCAGTCCCGGGTATATTCATTAAGAAAG 766
 |||||
 Db
 661 GCTTACCATTAAGAAACATCACTAAGACACCCAGTCCCGGGTATATTCATTAAGAAAG 720
 |||||
 QY
 767 AGAAGTCTGAGCTGAGAGAGAGGCTGTACCATTCATGCAACCCAGAGGGAATTCTG 826
 |||||
 Db
 721 AGAAGTCTGAGCTGAGAGAGGCTGTACCATTCATGCAACCCAGAGGGAATTCTG 780
 |||||
 QY
 827 AAGCATGCCGATGATCTTGAATTCATGACAGAAAGGAGAGATGAGCCAAAC-TAGCC 885
 |||||
 Db
 781 AAGCATGCCGATGATCTTGAATTCATGACAGAAAGGAGAGATGAGCCAAACTTAGCC 840
 |||||
 QY
 886 GAAGAGATTTC 895
 |||||
 Db
 841 GAAGAGATTTC 850

RESULT 7
 BX644668
 LOCUS
 DEFINITION
 DKFZP781G0234 r1 781 (synonym: h1cc4) Homo sapiens cDNA clone
 BX644668
 BX644668.1 GI:34479001
 EST.
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 832)
 Lauber, J., Bahr, A., Mewes, H.W., Weil, B., Amid, C., Osaenger, A.,
 Fodor, G., Han, M., and Wiemann, S.
 EST (Lauber, J., Bahr, A., Mewes, H.W., Weil, B., Amid, C., et al.)
 Unpublished (2003)
 Contact: MIPS
 MIPS
 Ingolstaedter Landstr.1, D-85764 Neuberg, Germany
 This is the 5' sequence of the clone insert.
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
 consortium of the German Genome Project.
 No si sequence available.
 This clone (DKFZP781G0234) is available at the RZPD in Berlin.
 Please contact the RZPD: Reesourcenzentrum, Heubenerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
 Location/Qualifiers
 1..832
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFZP781G0234"
 /dev stage="adult"
 /lab_host="DH10B"

ORIGIN /clone_lib="781 (synonym: hicc4)"
/note="Vector: pSport1_sfi; Site_1: SfiI; Site_2: SfiIb;
cDNA-collection"

Query Match 24.6%; Score 809; DB 13; Length 832;
Best Local Similarity 99.9%; Pred. No. 7.4e-148;
Matches 820; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
QY 1603 GAGGAAGTATGTCAGAAATTAATCGGGACTTTCTTGTACGACACTGCAACGGCAG 1662
DB 13 GAGGAAGTATGTCAGAAATTAATCGGGACTTTCTTGTACGACACTGCAACGGCAG 72
QY 1663 ATCAGGAAATTTGTAACAACAGGTAGAGCAAGAGCAAGAAATCCCTCAGAGAGTCC 1722
DB 73 ATCAGGAAATTTGTAACAACAGGTAGAGCAAGAGCAAGAAATCCCTCAGAGAGTCC 132
QY 1723 TCACAGCGCAGCAAGTGAAGTCTCCACAGGACCAAGCAAAACAGATGATGAGCC 1782
DB 133 TCACAGCGCAGCAAGTGAAGTCTCCACAGGACCAAGCAAAACAGATGATGAGCC 192
QY 1783 TTCCAAACCTGACAGATGAGCAACCAACGACCAAGATGCGAGCAACCAAGA 1842
DB 193 TTCCAAACCTGACAGATGAGCAACCAACGACCAAGATGCGAGCAACCAAGA 252
QY 1843 CCATCTGAGGAATGAGAAGTCTGCGAGAGCGCCAGAGGACTCTGCGAGGCGCTGAGAAC 1902
DB 253 CCATCTGAGGAATGAGAAGTCTGCGAGAGCGCCAGAGGACTCTGCGAGGCGCTGAGAAC 312
QY 1903 CCCAGGGGCGCAGAGAGGGCGGGAGAGTCAAGCCAGTTTGCAGAAACCAACGACCCCG 1962
DB 313 CCCAGGGGCGCAGAGAGGGCGGGAGAGTCAAGCCAGTTTGCAGAAACCAACGACCCCG 372
QY 1963 CTTCCGCGCGCGCGAGGCTTCTGAGGCTTCAAGCATTCACCTTCAACATCCACTCGGATC 2022
DB 373 CTTCCGCGCGCGCGAGGCTTCTGAGGCTTCAAGCATTCACCTTCAACATCCACTCGGATC 432
QY 2023 TCTCTGAACTCCCAAGAGCTATCCCTTTAGTGAATCAACATAGTGAAGCTGTTCA 2082
DB 433 TCTCTGAACTCCCAAGAGCTATCCCTTTAGTGAATCAACATAGTGAAGCTGTTCA 492
QY 2083 AAGCCAAAGAAATGCAACACCTTTTCTGTGGCAATCGTCTGTACATGTGTACA 2142
DB 493 AAGCCAAAGAAATGCAACACCTTTTCTGTGGCAATCGTCTGTACATGTGTACA 552
QY 2143 TATTAGAAAGGAAGATGTTAAGATATGTGCTGTGGTTACAAGGCTGTCAGCG 2202
DB 553 TATTAGAAAGGAAGATGTTAAGATATGTGCTGTGGTTACAAGGCTGTCAGCG 612
QY 2203 GTAATATATTTAGAAATATATATCAATTAACCTCAACTCAATTTTAAATCAAT 2262
DB 613 GTAATATATTTAGAAATATATATCAATTAACCTCAACTCAATTTTAAATCAAT 672
QY 2263 ATTAATTTTTTTTTTTTTTAAAGAGAAAGCAAGCTTTTCTAGACTTTAAAGATAAGT 2322
DB 673 ATTAATTTTTTTTTTTTTTAAAGAGAAAGCAAGCTTTTCTAGACTTTAAAGATAAGT 752
QY 2323 CTTTGGAGAGTCTCAAGGTGTAGAGAGAGCTTTAGAGCAACCCGACAAATTTCAACCA 2382
DB 733 CTTTGGAGAGTCTCAAGGTGTAGAGAGAGCTTTAGAGCAACCCGACAAATTTCAACCA 791
QY 2383 GAGGGAATCTCGTCGAGAGAGCACTCAAGGAGTCTGGA 2423
DB 792 GAGGGAATCTCGTCGAGAGAGCACTCAAGGAGTCTGGA 832
```

RESULT 8 836 bp mRNA linear EST 02-MAY-2002
LOCUS BQ212500
DEFINITION AGENCODE_7675661 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6095809
ACCESSION BQ212500
VERSION BQ212500.1 GI:20392782

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC/DCMP/DRP

cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.lnl.gov>
Plate: LLM13368 row: n column: 02
High quality sequence stop: 670.

FEATURES
source
1..836
Location/Qualifiers

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6095809"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 72"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

ORIGIN

Query Match 24.3%; Score 799.2; DB 13; Length 836;
Best Local Similarity 98.7%; Pred. No. 6.2e-146;
Matches 826; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

```
QY 2039 GAGGCTATCCCTTTAGTGAATCAATAGTGAAGAGCTTCAAGCAAGCAAAATGC 2098
DB 1 GAGGCTATCCCTTTAGTGAATCAATAGTGAAGAGCTTCAAGCAAGCAAAATGC 60
QY 2099 ACAACCTTTTCTGTGGCAAAATGCTCTGTACATGTGTACATATTAGAAAGGAGA 2158
DB 61 ACAACCTTTTCTGTGGCAAAATGCTCTGTACATGTGTACATATTAGAAAGGAGA 120
QY 2159 TGTTAAGATATGAGGCTGTGGTTACACAGGCTGCTGACGGTAAATATTATTTGAA 2218
DB 121 TGTTAAGATATGAGGCTGTGGTTACACAGGCTGCTGACGGTAAATATTATTTGAA 180
QY 2219 ATTAATATCAATTAATCACTCACTCAATTTTAAATCAATTTTATTTTCT 2278
DB 181 ATTAATATCAATTAATCACTCACTCAATTTTAAATCAATTTTATTTTCT 240
QY 2279 TTTTAAAGGAAAGCAGGCTTTCTAGACTTTAAAGATAAGTCTTGGAGGCTCAC 2338
DB 241 TTTTAAAGGAAAGCAGGCTTTCTAGACTTTAAAGATAAGTCTTGGAGGCTCAC 300
QY 2339 GGTGTAGAGAGAGCTTTGAGGCAACCCGCAAAATTTCAACGAGGAAATCTGTGC 2398
DB 301 GGTGTAGAGAGAGAGCTTTGAGGCAACCCGCAAAATTTCAACGAGGAAATCTGTGC 360
QY 2399 GAAGCACTCAAGCAGTCTGATCACTGTGTATGTCAACAGAGGATACGCTGC 2458
DB 361 GAAGCACTCAAGCAGTCTGATCACTGTGTATGTCAACAGAGGATACGCTGC 420
QY 2459 CTTGAAGAGAAATCTGTCACTCTCATGCTGTCTAGCTATACACCATTTCTCTT 2518
DB 421 CTTGAAGAGAAATCTGTCACTCTCATGCTGTCTAGCTATACACCATTTCTCTT 480
QY 2519 GCTCAAGGTTTAACTGTCTTTTGTGATCTGTATATATTTCTCTCTCTG 2578
```

Db 481 GCTTCACAGGTTTAAAGTGGTTTTCATCTGATGATATTAATTCCTGCTCTCTG 540
QY 2579 TTTATCTCTCCCTCCCTCCCTCCCTCTTCTTCATCTGATTTCTTTGAATTTCTC 2638
Db 541 TTTATCTCTCCCTCCCTCCCTCCCTCTTCTTCATCTGATTTCTTTGAATTTCTC 600
QY 2639 ATCCCTCATCTCAATCCCGATCTAGACACCCCGCCCGCCGAGGCAAGACAGTCTCT 2698
Db 601 ATCCCTCATCTCAATCCCGATCTAGACACCCCGCCCGCCGAGGCAAGACAGTCTCT 659
QY 2699 GAGTATCATCTACACAAAGAGAAACAAAGGAGAAACACAAACACAGCTCACTTAC 2758
Db 660 GAGTATCATCTACACAAAGAGAAACAAAGGAGAAACACAAACACAGCTCACTTAC 719
QY 2759 TTGGTACTCAAAAGAACAGTCAATGATCTTGTCTTA-GGTTTGGAGAGAGAA 2817
Db 720 TTGGTACTCAAAAGAACAGTCAATGATCTTGTCTTANCGTTTGGAGAGAGAA 779
QY 2818 ACAGGAACCCACCAACCAACCAATCAACCAACCAACCAACCAACCAATCAACCA 2874
Db 780 ACAGGAACCCACCAACCAACCAATCAACCAACCAACCAACCAACCAATCAACCA 836

RESULT 9
BM806132 1089 bp mRNA linear EST 05-MAR-2002
LOCUS BM806132
DEFINITION AGENCOURT 6553922 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5555652
5', mRNA sequence.
ACCESSION BM806132
VERSION BM806132.1 GI:19122955
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1089)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgs@bbs-rcmail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/HLN at:
http://image.llnl.gov
Plate: LHAM12376 row: c column: 13
High quality sequence stop: 688.

FEATURES
Source
1..1089
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5555652"
/issue_type="1e1omyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_71"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1: Not1;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."

ORIGIN
Query Match 24.3%; Score 797; DB 12; Length 1089;
Best Local Similarity 94.3%; Pred. No. 1.6e-145;
Matches 905; Conservative 0; Mismatches 41; Indels 14; Gaps 7;

QY 201 CTAACCCGACGAACTGGGCGCATCGCGCATCGAGACCCCTCGGGTAAATGGAAT 260
Db 1 CTAACCCGACGAACTGGGCGCATCGCGCATCGAGACCCCTCGGGTAAATGGAAT 60
QY 261 GATGGAAGAAATCATGGAATGATTAATCAATCTTAAAGAGAGAGAGAGAAAT 320

Db 61 GCATGGGAAATATCATGGAATGATTAATCAATCTTAAAGAGAGAGAGAGAAAT 120
QY 321 TCGAATTCGAAACATCCCTCCCTACCTGACAGTGGAGAGTGTGGATGACATTTGGCTCA 380
Db 121 TCGAATTCGAAACATCCCTCCCTACCTGACAGTGGAGAGTGTGGATGACATTTGGCTCA 180
QY 381 ATATGGACAGTGGAGATGTGGAACAATGACACACACACACACACACACACACACAC 440
Db 181 ATATGGACAGTGGAGATGTGGAACAATGACACACACACACACACACACACACACAC 240
QY 441 CGTACATATGCAACAGAGAAAGCAAAATGCCATGGAAGAGTGAAGCGGACATCA 500
Db 241 CGTACATATGCAACAGAGAAAGCAAAATGCCATGGAAGAGTGAAGCGGACATCA 300
QY 501 GTTGGAGAACTACTCTTCAAGATTTCTCAATCCCGATGAAGAGTGAAGCTCCCTTC 560
Db 301 GTTGGAGAACTACTCTTCAAGATTTCTCAATCCCGATGAAGAGTGAAGCTCCCTTC 360
QY 561 GCCCCTCAGGAGCCGAGCGTGGGACCACTCTTCCCGGAGACAGGCGCCCTGG 620
Db 361 GCCCCTCAGGAGCCGAGCGTGGGACCACTCTTCCCGGAGACAGGCGCCCTGG 420
QY 621 GGGCACTTCTAGGCGACAGATTTGATTTCCCGTGGGATCCCTGGCCCAACAGT 680
Db 421 GGGCACTTCTAGGCGACAGATTTGATTTCCCGTGGGATCCCTGGCCCAACAGT 480
QY 681 TGTGGTGCATCTACGGAAGAGAGGCTTGACATTAAGAACATCACTAAGCAGACCA 740
Db 481 TGTGGTGCATCTACGGAAGAGAGGCTTGACATTAAGAACATCACTAAGCAGACCA 540
QY 741 GTCCCGGTGATATTCATTAAGAAAGAGAACTCTGAGCTGACAGAAAGCTGTACAT 800
Db 541 GTCCCGGTGATATTCATTAAGAAAGAGAACTCTGAGCTGACAGAAAGCTGTACAT 600
QY 801 CCATGCGACCCGAGGAGGAGCTTGAAGACATGCGCATGATTTCTTGAATCATGACAGA 860
Db 601 CCATGCGACCCGAGGAGGAGCTTGAAGACATGCGCATGATTTCTTGAATCATGACAGA 660
QY 861 AGAGGCAATGAGACCAACTAGCCGAGAGATTCCTGTGAATCTTGGACACAGATG 920
Db 661 AGAGGCAATGAGACCAACTAGCCGAGAGATTCCTGTGAATCTTGGACACAGATG 720
QY 921 --CTTGTGGAAAGCTAATTTGAAAGAGGACAGAAATTTGAAGAAATTTGAACATGA 978
Db 721 GCTTGTGGAAAGCTAATTTGAAAGAGGACAGAAATTTGAAGAAATTTGAACATGA 780
QY 979 ACA-GGACCAAGATTAACATCTCTT-GCAGGATTTGACATATCAACCCGG-AA 1035
Db 781 ACAGGGACCAAGATTAACATCTCTT-GCAGGATTTGACATATCAACCCGGAA 840
QY 1036 AGAACCATCACTG-TGAAGGGACAG-TTGAAGGCTGTGCAAGTGTGAATAGATTA 1093
Db 841 AGAACCATCACTGTTGAAGGGACAGTTGAGCCCTGTGCCATGCTGAATAGATTA 900
QY 1094 TGAAGAGCTG-----GTAGAGCTTTGAAATGATATCTGCTTTAAACACCCAC 1146
Db 901 TATGAAAGAACTGTGCGCGAGAGCCCTTTTGAAGATGAATGTGGCTGTTTAAACAA 960

RESULT 10
BU156245 894 bp mRNA linear EST 03-SEP-2002
LOCUS BU156245
DEFINITION AGENCOURT 7965128 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6162660
5', mRNA sequence.
ACCESSION BU156245
VERSION BU156245.1 GI:22669777
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 894)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
Plate: L1AM13516 row: 0 column: 13
High quality sequence stop: 666.
Location/Qualifiers

FEATURES
source 1..894
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6162660"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_72"
/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

ORIGIN

Query Match 23.1%; Score 759.8; DB 13; Length 894;
Best Local Similarity 95.5%; Pred. No. 3.2e-138;
Matches 836; Conservative 0; Mismatches 32; Indels 7; Gaps 5;

```
QY 1973 CCCAGGGCTTCGAGGCTTCAGCATCCATTCACATCCACTCGGATCTCTCTGAAC 2032
DB 20 CCTCTCTACCTCGAGGCTTCAGCATCCATTCACATCCACTCGGATCTCTCTGAAC 79
QY 2033 TCCCAGAGCGCTATCCCTTTTGAAGTGAATACATAGTGAACGCTTCAAGCCAGCA 2092
DB 80 TCCCAGAGCGCTATCCCTTTTGAAGTGAATACATAGTGAACGCTTCAAGCCAGCA 139
QY 2093 AAATGCAACCCCTTTTCTGTGCGAAATGCTCTGTACATGTGTACATATTAGAAAG 2152
DB 140 AAATGCAACCCCTTTTCTGTGCGAAATGCTCTGTACATGTGTACATATTAGAAAG 139
QY 2153 GGAAGATGTTAATATGAGGCGCTGTGAGTTACACAGGGTGCTCGACGGCTAATATAT 2212
DB 200 GGAAGATGTTAATATGAGGCGCTGTGAGTTACACAGGGTGCTCGACGGCTAATATAT 259
QY 2213 TTAGAATATATATATCAATATCACTCACTCACTCACTCACTCACTCACTCACTCACT 2272
DB 260 TTAGAATATATATATCAATATCACTCACTCACTCACTCACTCACTCACTCACTCACT 319
QY 2273 TTTTCTTTTAAAGAGAAAGCAGGCTTTCTAGACTTTTAAAGATTAAGCTTTGGGAGG 2332
DB 320 TTTTCTTTTAAAGAGAAAGCAGGCTTTCTAGACTTTTAAAGATTAAGCTTTGGGAGG 379
QY 2333 TGTCAAGGTGTGAGAGGAGCTTTGAGGCCACCCCGCAAAATTCACCCAGAGGAAATC 2392
DB 380 TGTCAAGGTGTGAGAGGAGCTTTGAGGCCACCCCGCAAAATTCACCCAGAGGAAATC 439
QY 2393 TGTGCGAAGACACTCAGGCGAGTCTGTGATCACTGTGATGTCAACAGAAAGGATTC 2452
DB 440 TGTGCGAAGACACTCAGGCGAGTCTGTGATCACTGTGATGTCAACAGAAAGGATTC 499
QY 2453 CGTCTCTTGAAGAGAAACTCTGTCACTCTCATCTCTAGCTATACACCAATTT 2512
DB 500 CGTCTCTTGAAGAGAAACTCTGTCACTCTCATCTCTAGCTATACACCAATTT 559
QY 2513 CTCTTGTCTTCAAGGTTTAAAGCTTTTGTGATCTGCTATATTAATTTCTGTCTC 2572
DB 560 CTCTTGTCTTCAAGGTTTAAAGCTTTTGTGATCTGCTATATTAATTTCTGTCTC 619
```

```
QY 2573 TCTCGTTTATCTCTCCCTCCCTCCCTCCCTCCCTCTCTCTCTCATCTGCA-TTCTTTTGA 2631
DB 620 TCTCTGTATCTCTCCCTCCCTCCCTCCCTCCCTCTCTCTCTCATCTCTCTTTTGA 679
QY 2632 TTTCTCATCCCTCATCTCAATCCCGATCTATGACACCCCGCCCCCGGAGCAAGCA 2691
DB 680 TTTCTCATCCCTCATCTCAATCCCGATCTATGACCA-CCCCCCCCCGGAGCAAGCA 737
QY 2692 GTGCTGTAGTATCATATCACAACAAAGGAAACAAAGGCAACACAAACAGCTTCA 2751
DB 738 GTGCTGTAGTATCATATCACCACCAAGGAAACAAAGGCAACACAAACAGCTTCA 797
QY 2752 C-TTACACTTGCTATCTCAAAAG-AAACAAGTGAATGCTTCTGTAGCTTTGG- 2808
DB 798 CTTTACACTTGCTATCTCAAAAGAAACAAAGGCAAGGCTTATCTGTAGCTTTGG 857
QY 2809 -AAGAGAAAACAGAAACCCACCAACCAACCAAT 2842
DB 858 AAGAGAAAACAGAAACCCACCAACCAACCAAT 892
```

RESULT 11
BG575889 781 bp mRNA linear EST 10-APR-2001
LOCUS BG575889
DEFINITION 602598315F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4707243 5',
mRNA sequence.
ACCESSION BG575889
VERSION BG575889.1 GI:13583542
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 781)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
Plate: L1AM10576 row: m column: 04
High quality sequence stop: 773.
Location/Qualifiers

FEATURES
source 1..781
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4707243"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_87"
/note="Organ: breast; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."

ORIGIN

Query Match 22.9%; Score 752.6; DB 12; Length 781;
Best Local Similarity 98.6%; Pred. No. 8.3e-137;
Matches 770; Conservative 0; Mismatches 9; Indels 2; Gaps 1;

```
QY 1692 GCAGAGACGAAATACCTCAGGAGTGCCTCAGACGCCAGCAGTGAAGGCTCCCAAG 1751
DB 1 GCAGAGACGAAATACCTCAGGAGTGCCTCAGACGCCAGCAGTGAAGGCTCCCAAG 60
QY 1752 GCACACGAAAACAGATGATGATGAGCCCTTCCACACCTGACAGAAATGAGACCAAC 1811
```

```

Db      61  GCACCCAGCAAAACAGGATGATGATGTCTCTTCCAAACCTGACAGAAATGAGCCAAAC 120
QY      1812 GCAGCCAGCAGATCGGAGGCAAAACCAAGACATCTGAGGATGAGAAGCTCTCGGAGG 1871
Db      121  GCAGCCAGCAGATCGGAGGCAAAACCAAGACATCTGAGGATGAGAAGCTCTCGGAGG 180
QY      1872 CGGCCAGGAGACTCTGCCGAGGCTCTGAGAACCCCAAGGCGGAGAGGCGGAGGAGGT 1931
Db      181  CGGCCAGGAGACTCTGCCGAGGCTCTGAGAACCCCAAGGCGGAGAGGCGGAGGAGGT 240
QY      1932 CAGCCAGGTTGGCAGAAACCAACGAGCCCGCTCCGCGCCCGCAGGCTCTCTGAGGCT 1991
Db      241  CAGCCAGGTTGGCAGAAACCAACGAGCCCGCTCCGCGCCCGCAGGCTCTCTGAGGCT 300
QY      1992 TCAGGCATCCACTTCCACCATCCACTCGGATCTCTCTGAGTCTCCGAGCTCCGAGCTATCCCTT 2051
Db      301  TCAGGCATCCACTTCCACCATCCACTCGGATCTCTCTGAGTCTCCGAGCTATCCCTT 360
QY      2052 TTAGCTTGAACCTAACCTAGGTAAGCTGTTCAAAAGCCAGCAAAATGACACCTTTTCT 2111
Db      361  TTAGCTTGAACCTAACCTAGGTAAGCTGTTCAAAAGCCAGCAAAATGACACCTTTTCT 420
QY      2112 GTGGCAATTCGCTCTGTACATGTGTGTACATATTAGAAAGGAGATGTTAAGATATGT 2171
Db      421  GTGGCAATTCGCTCTGTACATGTGTGTGTACATATTAGAAAGGAGATGTTAAGATATGT 480
QY      2172 GAGCTGTGGGTTACACAGGAGTCCCTGACGCGGTAAATATTTTGTAGAAATATATTCAAA 2231
Db      481  GAGCTGTGGGTTACACAGGAGTCCCTGACGCGGTAAATATTTTGTAGAAATATATTCAAA 540
QY      2232 TTAACCACTAACCTCAATTTTAAATTAATTAATTTTCTTTTAAAGAGAA 2291
Db      541  TTAACCACTAACCTCAATTTTAAATTAATTAATTTTCTTTTAAAGAGAA 600
QY      2292 GCAGGCTTTTCTAGACTTTAAAGAAATTAAGTCTTTGAGAGTCTCAAGTGTAGAGAGA 2351
Db      601  GCAGGCTTTTCTAGACTTTAAAGAAATTAAGTCTTTGAGAGTCTCAAGTGTAGAGAGA 660
QY      2352 GCTTTGAGGAGCAGCCGACAAATTCACCAAGAGGAAATCTCTGTGAGAGACATCTAC 2411
Db      661  GCTTTGAGGAGCAGCCGACAAATTCACCAAGAGGAAATCTCTGTGAGAGACATCTAC 720
QY      2412 --GGCAGTCTGAGTACCTGTGTATGTGACAGAGAGGATACCGTCTCTTGAAGAGA 2469
Db      721  CGCAGTCTTCTGATCACTGTGTATGTGACAGAGAGGATACCGTCTCTTGAAGAGA 780
QY      2470 A 2470
Db      781 A 781

```

```

RESULT 12
CD243351 940 bp mRNA linear EST 22-MAY-2003
LOCUS AGENCOURT_14121251 NIH_MGC_180 Homo sapiens cDNA clone
DEFINITION IMAGE:30383283 5', mRNA sequence.
ACCESSION CD243351
VERSION CD243351.1 GI:31003815
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 940)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-f@mail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)

```

DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LIML at:
<http://image.liml.gov>
 Plate: NDM446 row: 0 column: 04
 High quality sequence start: 18
 High quality sequence stop: 695.
 Location/Qualifiers
 1. 940
 /organism="Homo sapiens"
 /mol_type="mRNA"

FEATURES
 source
 /db_xref="taxon:9606"
 /clone="IMAGE:30383283"
 /lab_host="DH10B-Ton A (T1 and T5 phage resistances)"
 /clone_id="NIH_MGC_180"
 /note="Organ: Testis; Vector: pCMV-Sport6.1; Site: 1; NotI;
 Site 2: EcoRV (destroyed); Library is oligo-dT primed and
 directionally cloned (EcoRV site is destroyed upon
 cloning). Average insert size 1.68 kb. Library was
 constructed by (Invitrogen). Note: this is a NIH-MGC
 Library."

ORIGIN

Query Match 22.9%; Score 751.4; DB 14; Length 940;
 Best Local Similarity 99.5%; Pred. No. 1.4e-136; Indels 3; Gaps 3;
 Matches 785; Conservative 0; Mismatches 1;

```

QY      125  TCCGCGACCTCTTTGGGAGCAGAGAGCTGCCCTGGCGGAGCAGGCTCTGTAAGTCCG 184
Db      28  TCCGCGACCTCTTTGGGAGCAGAGAGCTGCCCTGGCGGAGCAGGCTCTGTAAGTCCG 87
QY      185  GCTACGCTTGTGTGACTACCCCGACAGAACTGGGCGCATCGGCCATCCGAGACCTCT 244
Db      88  GCTACGCTTGTGTGACTACCCCGACAGAACTGGGCGCATCGGCCATCCGAGACCTCT 147
QY      245  CCGGTAAAGTGGAAATTCATGAGGAAATTCATGAGGATGATTACTCACTCTTAAAGC 304
Db      148  CCGGTAAAGTGGAAATTCATGAGGAAATTCATGAGGATGATTACTCACTCTTAAAGC 207
QY      305  TAAGGAGCAGGAAATTCAGATTCGAAACATCCCTCTCACTGAGTGGAGGCTTGG 364
Db      208  TAAGGAGCAGGAAATTCAGATTCGAAACATCCCTCTCACTGAGTGGAGGCTTGG 267
QY      365  ATGACTTTTGGCTCAATATGAGGACATGTGAGAAATGTGAAACAAGTCAACAGACAG 424
Db      268  ATGACTTTTGGCTCAATATGAGGACATGTGAGAAATGTGAAACAAGTCAACAGACAG 327
QY      425  AAACCGCCCTGTCTCAACGTCAATATGCAACAAGAGAAAGCAAAATAGCCATGAGA 484
Db      328  AAACCGCCCTGTCTCAACGTCAATATGCAACAAGAGAAAGCAAAATAGCCATGAGA 387
QY      485  AGCTAAGCGGAGATCAGTTTGAGAACTACTCTTCAGAAATTTCTTACATCCGGATGAG 544
Db      388  AGCTAAGCGGAGATCAGTTTGAGAACTACTCTTCAGAAATTTCTTACATCCGGATGAG 447
QY      545  AGGTGAGCTCCCTTGTGCCCCCTCAGCGAGCCGAGTGGGAGCCCTCCCGGAGAG 604
Db      448  AGGTGAGCTCCCTTGTGCCCCCTCAGCGAGCCGAGTGGGAGCCCTCCCGGAGAG 507
QY      605  AAGGCCAGCCCTCGGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCGCTGGGATCC 664
Db      508  AAGGCCAGCCCTCGGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCGCTGGGATCC 567
QY      665  TGGTCCCAACCCAGTTTGTGTGTCATCATCGGAAAGAGAGGCTTGACATTAAGACA 724
Db      568  TGGTCCCAACCCAGTTTGTGTGTCATCATCGGAAAGAGAGGCTTGACATTAAGACA 627
QY      725  TCATTAAGACACCATCCCGGTGATATCTATGAAAAAGAACTCTGAGGCTGAG 784
Db      628  TCATTAAGACACCATCCCGGTGATATCTATGAAAAAGAACTCTGAGGCTGAG 687
QY      785  AGAAGCTGTCAACATCCCAACCCAGAGGGGACTTCTGAAGCATGCGCATGATTC 844

```

Db	688	AGAAAGCCCTGTACACATTCATCCATGCAACCCAGAGGGAGCTTCTGAGATGATCCGATATTC	747
QY	845	TTGAAATTCATTCGA-GAAAGAGCGCATGTAGACCAACTAG-CCGAGAGATT-CCCTCTGA	901
Db	748	TTGAAATTCATTCGAGGAGAGAGCGATGAGACCAACTAGCCGAGAGATTCCTCTGA	807
QY	902	AAATCTTG 910	
Db	808	AAATCTTG 916	
RESULT 13			
LOCUS	CF593505		
DEFINITION	CF593505	824 bp	mRNA linear EST 26-SEP-2001
ACCESSION	AGENCOURT_15622249 NIH MGC_147 Homo sapiens cDNA clone		
VERSION	IMAGE:30531076 5', mRNA sequence.		
KEYWORDS	CF593505		
SOURCE	CF593505.1 GI:36347108		
ORGANISM	EST.		
REFERENCE	Homo sapiens (human)		
AUTHORS	Homo sapiens		
TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
JOURNAL	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
COMMENT	1 (bases 1 to 824)		
	NIH-MGC http://mgi.nci.nih.gov/		
	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Daniela S. Gerhard, Ph.D.		
	Office of Cancer Genomics		
	National Cancer Institute / NIH		
	Bldg. 31 Rm10A07 Bethesda, MD 20892		
	Email: cgapbs-rt@mail.nih.gov		
	Tissue Procurement: Dr. Stefan Hanson		
	cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help		
	and advice from Piero Carninci (RIKEN)		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNL at:		
	http://image.lnl.gov		
	Plate: NDAM621 row: e column: 05		
	High quality sequence stop: 682.		
FEATURES	Location/Qualifiers		
SOURCE	1..824		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:30531076"		
	/tissue_type="Human Placenta"		
	/lab_host="DH10B Tona"		
	/clone_lib="NIH MGC 147"		
	/note="Organ: placenta; Vector: pBluescript; Site_1:		
	all-XhoI; Site_2: BamH; Oligo-dT primed using primer		
	5'-TTTTTTTTTTTTTTT-3', size-selected for average		
	insert size 2.3 kb and normalized to ROT 5. This is a		
	primary library enriched for full-length clones and		
	constructed using the Cap-trapper method (carninci, in		
	preparation). Library constructed by M. Brownstein		
	(NIH/NHGRI, National Institutes of Health). Note: This is		
	a NIH_MGC library."		
ORIGIN			
Query Match	22.8%; Score 749.8; DB 14; Length 824;		
Best Local Similarity	99.4%; Pred. No. 2.9e-116;		
Matches 773; Conservative 0; Mismatches 3; Indels 2; Gaps 2;			
QY	1	GGCAGCGGAGAGCGCGCGGATCCGCGGAGCGCGGAGCGCGGCTCTCGGG 60	
Db	38	GGCAGCGGAGAGCGCGGATCCGCGGATCCGCGGAGCGCGGAGCGCGGCTCTCGGG 97	
QY	61	AAGAGCGGATGATGAAACAAGCTTTACATCGGGAACCTGAACTCCCGCTGACCGCGAC 120	
Db	98	AAGAGCGGATGATGAAACAAGCTTTACATCGGGAACCTGAACTCCCGCTGACCGCGAC 157	

QY	121	GACCTCCGGCACTCTTTGGGGACAGAAAGCTGCCCTTGCGGACAGCTCCGCTGAAG	180
Db	158	GACCTCCGGCACTCTTTGGGGACAGAAAGCTGCCCTTGCGGACAGCTCCGCTGAAG	217
QY	181	TCGGGCTACGCCCTGTGTGACTACCCGACCAAACTGGGACATCGCCCATTCAGACC	240
Db	218	TCGGGCTACGCCCTGTGTGACTACCCGACCAAACTGGGACATCGCCCATTCAGACC	277
QY	241	CTCTGGGCTAAAGTGGAAATTGCATGGGAAATCATGGAACTTGAATTAATCTCTTAA	300
Db	278	CTCTGGGCTAAAGTGGAAATTGCATGGGAAATCATGGAACTTGAATTAATCTCTTAA	337
QY	301	AAGCTAAGAGCAGGAAAAATTCAGATTGGAAATCCCTCCATCCCTGACAGTGGAGG	360
Db	338	AAGCTAAGAGCAGGAAAAATTCAGATTGGAAATCCCTCCATCCCTGACAGTGGAGG	397
QY	361	TTGATGAGCTTTTGGCTCAATATATGAGCAAGTGGAAATGTGGAACTAGTCAACAGAC	420
Db	398	TTGATGAGCTTTTGGCTCAATATATGAGCAAGTGGAAATGTGGAACTAGTCAACAGAC	457
QY	421	AACGAAACCGCCGTTGTCAAGCTCACTATGCAACAAGAAAGCAAAAATAGCCATG	480
Db	458	AACGAAACCGCCGTTGTCAAGCTCACTATGCAACAAGAAAGCAAAAATAGCCATG	517
QY	481	GAGAAGCTAAGCGGACATCACTTGAAGAATCTCTGCTCAAGATTCTCAATCCCGAT	540
Db	518	GAGAAGCTAAGCGGACATCACTTGAAGAATCTCTGCTCAAGATTCTCAATCCCGAT	577
QY	541	GAAGAAGTGAAGCTCCCTTCCTGCGCCCTCAGCGAGGCCAGCGTGGGAGACACTTCCCGG	600
Db	578	GAAGAAGTGAAGCTCCCTTCCTGCGCCCTCAGCGAGGCCAGCGTGGGAGACACTTCCCGG	637
QY	601	GAGCAAGGCTACGCCCCCTTGGGGGCACTTCTCAGGCTCAACAGATTGATTTCCCGTGG	660
Db	638	GAGCAAGGCTACGCCCCCTTGGGGGCACTTCTCAGGCTCAACAGATTGATTTCCCGTGG	697
QY	661	ATCCGATGCCCAACCCAGTTTGTGTGTGCCATCATCGGAAAGAGGCGCTTACCACTAAG	720
Db	698	ATCCGATGCCCAACCCAGTTTGTGTGTGCCATCATCGGAAAGAGGCGCTTACCACTAAG	757
QY	721	AACATCACTAAGCAGACCAGTCCG-GGGTGAATATCATTA-GAAAGAGAACTGTGG	776
Db	758	AACATCACTAAGCAGACCAGTCCG-GGGTGAATATCATTA-GAAAGAGAACTGTGG	815
RESULT 14			
LOCUS	BG748346	889 bp	mRNA linear EST 15-MAY-2001
DEFINITION	6027050221 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4842553 5',		
ACCESSION	BG748346		
VERSION	BG748346.1	GI:14058999	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
TITLE	1 (bases 1 to 889)		
JOURNAL	NIH-MGC http://ngc.nci.nih.gov/ .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabs@remail.nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Ling Hong/Rubin Laboratory		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLML)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNLML at:		
	http://image.lnl.gov		
	Plate: LNCM1677 row: c column: 02		
	High quality sequence stop: 773.		

FEATURES

Source

Location/Qualifiers

1..889
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:484253"
 /tissue_type="normal pigmented retinal epithelium"
 /lab_host="DH10B (phage-resistant)"
 /clone_1lb="NIH_MGC_43"
 /note="Organ: eye; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library." "

ORIGIN

Query Match 22.7%; Score 746; DB 12; Length 889;
 Best Local Similarity 95.5%; Pred. No. 1.6e-135;
 Matches 802; Conservative 0; Mismatches 30; Indels 8; Gaps 3;

```

72 GATGAACAAGCTTTTACATCGGGAACTGAGCCCGCGTCAACCGCGAGACCTCCGGCA 131
2 GATGAACAAGCTTTTACATCGGGAACTGAGCCCGCGTCAACCGCGAGACCTCCGGCA 61
132 GCTCTTTGGGAGCAGAAAGCTCCCTGGCGGAGCAAGTCTCTGTAAGTCCGGCTACGC 191
62 GCTCTTTGGGAGCAGAAAGCTCCCTGGCGGAGCAAGTCTCTGTAAGTCCGGCTACGC 121
132 CTTCTGTGACCTACCCGACCAAGCTCCGCGGCATCCGCGCATCGAGACCTCCGGCTAA 251
122 CTTCTGTGACCTACCCGACCAAGCTCCGCGGCATCCGCGCATCGAGACCTCCGGCTAA 181
252 AGTGAATTCATCGGGAAATCATGGAAAGTTGATTAAGTCTCTAATAAGCTAAGAG 311
132 AGTGAATTCATCGGGAAATCATGGAAAGTTGATTAAGTCTCTAATAAGCTAAGAG 241
312 CAGGAAATTCATCGGGAAATCATGGAAAGTTGATTAAGTCTCTAATAAGCTAAGAG 371
242 CAGGAAATTCATCGGGAAATCATGGAAAGTTGATTAAGTCTCTAATAAGCTAAGAG 301
372 TTTGGCTCAATTTGGAGAGTGGAGAAATGGAGAACAGTCAACAGACAGACAGAAACGGC 431
302 TTTGGCTCAATTTGGAGAGTGGAGAAATGGAGAACAGTCAACAGACAGACAGAAACGGC 361
432 CTTGTCAACGTCAATATGCAACAGAGAGAGCAAAATATAGCCATGAGAGCTAAG 491
362 CTTGTCAACGTCAATATGCAACAGAGAGAGCAAAATATAGCCATGAGAGCTAAG 421
492 CGGCGATCATGTTGAGAACTACTCTCTTCAAGATTTCTTACATCCCGATGAGAGGTGAG 551
422 CGGCGATCATGTTGAGAACTACTCTCTTCAAGATTTCTTACATCCCGATGAGAGGTGAG 481
552 CTCCCTTTGGCCCTCAGAGAGCCAGAGGTGGGAGCAACTCTTCCCGGAGAGAGGCA 611
482 CTCCCTTTGGCCCTCAGAGAGCCAGAGGTGGGAGCAACTCTTCCCGGAGAGAGGCA 541
612 CGCCCTCGGGGCACTTCTCAGGCGAGACAGATTTGATTTCCGCTCGGATCTCTGTC 671
542 CGCCCTCGGGGCACTTCTCAGGCGAGACAGATTTGATTTCCGCTCGGATCTCTGTC 601
672 CACCGATTTGTTGTGTCATCATCGAAAAGAGGCGTTGACCATTAAGAACTCACTAA 731
732 GAGAGACCGATCCCGGATGATTCATGAGAAAGAAAGTCTGAGCTGACAGAGAGC 791
662 GAGAGACCGATCCCGGATGATTCATGAGAAAGAAAGTCTGAGCTGACAGAGAG-C 720
792 TGTCAACATCATGACACCCAGAGGAGGAGCTTCTGAA--GCATGCGCATGATCTTGA 849

```

Db 721 TGTCAACATCATGACACCCAGAGGAGGAGCTTTAAAGCATTTGCCCATGATTTCTGTA 780
 Oy 850 ATCATGCAAGAAAGAGC-----AGATGAGACCAACTAGCCGAGAGATTCCTTGAAA 904
 Db 781 AATCTGGACAGACAGCCGATTTTGAACCAAAATAGCCGAGAAATGCTCTGTTAA 840

RESULT 15
 BU601634
 LOCUS
 DEFINITION
 IMAGE:6495584 5', mRNA sequence.
 ACCESSION
 BU601634
 VERSION
 BU601634.1 GI:23253393
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 762)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1998)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: NCI
 cDNA Library Preparation: Michael Brownstein Laboratory
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCN2673 row: k column: 09
 High quality sequence stop: 591.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

FEATURES

Source

Location/Qualifiers

1..762
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6495584"
 /tissue_type="mixed (pool of 40 RNAs)"
 /lab_host="DH10B (T1-phage-resistant)"
 /clone_1lb="NIH_MGC_142"
 /note="Vector: pDNR-LIB; Site 1: SfiI (ggccatgagcc); Site 2: SfiI (ggcgccctggcc); Double-stranded cDNA was prepared from a pool of 40 cell line polyA+ RNAs (bladder 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon - 4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%, kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%, ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were used in cloning as follows:
 5'-AAGCAGTGGTATCAACGAGAGTGGCGCATTAAGCGCGG-3' and
 5'-ATTCTAGAGCCGAGGCGGCGCCACAGG-dt(30)NN-3'. Pull-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the >0.5 kb size fraction (other fractions present in NIH_MGC 141). Library created in the laboratory of M. Brownstein (NIH, NIH). Note: this is a NIH_MGC library." "

ORIGIN

Query Match 21.9%; Score 718.8; DB 13; Length 762;
 Best Local Similarity 99.6%; Pred. No. 3.5e-130;
 Matches 731; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```

Oy 1 GCGAGCGAGAGAGCGCGCGGCTACCGCGCGCGGAGCGCGCGGCTCTCGGGG 60
Db 7 GCGAGCGAGAGAGCGCGCGGCTACCGCGCGCGGAGCGCGCGGCTCTCGGGG 66
Oy 61 AAGAGCGAGATGTAACAAGCTTTTACATCGGGAACCTGAGCCCGCGTCAACCGCGAC 120
Db 67 AAGAGCGAGATGTAACAAGCTTTTACATCGGGAACCTGAGCCCGCGTCAACCGCGAC 126

```



```
QY 121 GACCTCCGGCAGCTCTTTGGGACAGAAAGCTGCCCTTGCGGACAGATCCTGTGAAG 180
|
Db 127 GACCTCCGGCAGCTCTTTGGGACAGAAAGCTGCCCTTGCGGACAGATCCTGTGAAG 186
|
QY 181 TCCGGCTACGCTTCGTGACTACCCGACAGAACTGGGCCATCCGGCCATCGAGACC 240
|
Db 187 TCCGGCTACGCTTCGTGACTACCCGACAGAACTGGGCCATCCGGCCATCGAGACC 246
|
QY 241 CTCTCGGGTAAAGTGAATGCAATGGGAAATCAATGAAAGTTGATTAATCTAGTCTTAA 300
|
Db 247 CTCTCGGGTAAAGTGAATGCAATGGGAAATCAATGAAAGTTGATTAATCTAGTCTTAA 306
|
QY 301 AAGCTAAGAGCAGAGAAATTCAGATTCGAAACATCCCTCACTGACGTGGAGAGTG 360
|
Db 307 AAGCTAAGAGCAGAGAAATTCAGATTCGAAACATCCCTCACTGACGTGGAGAGTG 366
|
QY 361 TTGGATGACTTTTGGCTCAATATGGAACAGTGGAGATGTGAACAAGTCAACACAGAC 420
|
Db 367 TTGGATGACTTTTGGCTCAATATGGAACAGTGGAGATGTGAACAAGTCAACACAGAC 426
|
QY 421 ACAGAAACCGCCGTGTCACAGTCACATATGCAACAAGAGAGCAAAAATATGCCATG 480
|
Db 427 ACAGAAACCGCCGTGTCACAGTCACATATGCAACAAGAGAGCAAAAATATGCCATG 486
|
QY 481 GAGAGCTAAGCGGGCATCAAGTTGAGAACTACTCTTCAAGATTTCTAATCCCGGAT 540
|
Db 487 GAGAGCTAAGCGGGCATCAAGTTGAGAACTACTCTTCAAGATTTCTAATCCCGGAT 546
|
QY 541 GAAGAGGTGAGCTCCCTTGCCCTCAGCGAGCCCAAGCTGGGAGCAACTTCCCGG 600
|
Db 547 GAAGAGGTGAGCTCCCTTGCCCTCAGCGAGCCCAAGCTGGGAGCAACTTCCCGG 606
|
QY 601 GAGCAAGGCCAGC-CGCCGTGGGGCACTTCTCAGGCCAGACAGATTTTCCCGCTGCG 659
|
Db 607 GAGCAAGGCCAGCAGCCCTTG333CACTTCTCAAGCCAGACAGATTTTCCCGCTGCG 666
|
QY 660 GATCCTGATCCCAACCAAGTTGTGGTGCATCATCGAAAGAGAGGCTTGAACATAAA 719
|
Db 667 GATCCTGATCCCAACCAAGTTGTGGTGCATCATCGAAAGAGGCTTGAACATAAA 726
|
QY 720 GAACATCACTAAGC 733
|
Db 727 GAACATCACTAAGC 740
|
```

Search completed: July 24, 2004, 16:42:45
Job time : 8228.67 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 23, 2004, 10:51:25 ; Search time 144.09 Seconds

(without alignments)
12875.302 Million cell updates/sec

Title: US-09-270-437D-8
Perfect score: 5886
Sequence: 1 ggcagcgagagcgagcgagga.....aaccttgaatgttattt 3283

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+np_model -DEV=xlp
-O=/cgn2_1/USPTO.spool.p/US09270437/runat.23072004.095738.27403/app.query.fasta_1.5582
-DB-A Geneseq.29Jan04 -QPMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCT=0
-LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MTN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NOEM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US09270437.qcgn.1.185.qrunat.23072004.095738.27403 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A_Geneseq.29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2949	50.1	594	4	ABG06795
2	2838	48.2	556	5	ABG96346
3	2836.5	48.2	614	4	ABG06794
4	2808	47.7	620	6	ABU16163
5	2808	47.7	620	6	ABU55232
6	2584.5	43.9	555	6	ABU88799
7	1934	32.9	577	2	AAI30649
8	1875	31.9	579	3	AAI1365
9	1875	31.9	579	5	ABB75053
10	1875	31.9	579	5	ABB74997

11	1875	31.9	579	5	ABB75054	Abb75054 Human lun
12	1875	31.9	579	5	ABP61917	Abp61917 Human lun
13	1875	31.9	579	5	ABP61974	Abp61974 Human lun
14	1875	31.9	579	5	ABP61973	Abp61973 Human lun
15	1875	31.9	579	7	ADA28536	Ada28536 Recombina
16	1875	31.9	579	7	ADA28539	Ada28539 Recombina
17	1875	31.9	579	7	ADA28438	Ada28438 Human lun
18	1870	31.8	586	5	ABB75048	Abb75048 Human lun
19	1870	31.8	586	5	ABP61968	Abp61968 Human lun
20	1870	31.8	586	7	ADA28517	Ada28517 Recombina
21	1868	31.7	579	3	AAI13328	AAI13328 Human lun
22	1868	31.7	579	5	ABB74960	Abb74960 Human lun
23	1868	31.7	579	5	ABP61880	Abp61880 Human lun
24	1868	31.7	579	7	ADA28266	Ada28266 Human lun
25	1868	31.7	579	7	ADD14066	Add14066 Human src
26	1868	31.7	579	7	ADE53471	Ades53471 Human lun
27	1798	30.5	619	4	ABG21963	Abg21963 Novel hum
28	1725	29.3	583	4	ABG12592	Abg12592 Novel hum
29	1075.5	18.3	319	4	AAI93826	AAI93826 Human pol
30	872	14.8	187	6	AAO23971	AAO23971 Human IGF
31	855	14.5	558	4	ABB58367	Abb58367 Drosophil
32	812.5	13.8	261	4	AAU16161	AAU16161 Human nov
33	812.5	13.8	261	6	ABU55230	Abu55230 Human nov
34	803	13.6	171	4	AAU16166	AAU16166 Human nov
35	803	13.6	171	6	ABU55235	Abu55235 Human nov
36	782.5	13.3	250	4	AAU16579	AAU16579 Human nov
37	782.5	13.3	250	6	ABU55648	Abu55648 Human nov
38	750	12.7	171	4	AAU16583	AAU16583 Human nov
39	750	12.7	171	6	ABU55652	Abu55652 Human nov
40	676.5	11.5	192	4	AAU16164	AAU16164 Human nov
41	676.5	11.5	192	6	ABU55233	Abu55233 Human nov
42	501	8.5	192	3	AAI303261	AAI303261 Human sec
43	473.5	8.0	266	4	ABG12593	Abg12593 Novel hum
44	469.5	8.0	209	4	ABG21961	Abg21961 Novel hum
45	458	7.8	148	4	ABG21962	Abg21962 Novel hum

ALIGNMENTS

RESULT 1
ABG06795
ID ABG06795 standard; protein; 594 AA.

AC ABG06795;
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #6786.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
PI Dmanac RT, Liu C, Tang YT;
XX
DR WPI, 2001-639362/73.
DR N-PSDB; AAS70982.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess

PT biodiversity.
XX
PS Claim 20: SEQ ID NO 37154; 103bp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published/pct/sequences](http://wipo.int/pub/published/pct/sequences)

Sequence 594 AA;

Alignment Scores:	
Pred. No.:	4.87e-265
Score:	2949.00
Percent Similarity:	100.00%
Best Local Similarity:	99.83%
Query Match:	50.10%
Length:	594
Matches:	577
Conservative:	1
Mismatches:	0
Indels:	0

US-09-270-437D-8 (1-3283) X ABG06795 (1-594)

QY 4 AGCGGAGAGGCGAGGAGCGCGGGTACCGGCGCGGAGGAGCGCGGGGCTCTCTGGGGAAAG 63
 Db 17 AenglyGIyGIyGIuGIuAArgValProGIyAArgGISerAArgGlyIleuSerGIyIleuSerGIyLys 36
 QY 64 AAGCGATATGATAACAAAGCTTTACATCGGGAAACCTAGAGCCCGCGATCACCGCGAGCGAG 123
 Db 37 AArgArgMetMetCAsnLysIleuYrIIeGIyAsnLeuSerProAlIValThrAlaAspAsp 56
 QY 124 CTCGGGCACTCTCTTGGGGGACAGAAAGCTGCCCTGAGCGGACAGAGTCTCGTGAAGTCC 183
 Db 57 LeuAArgIleuPheGlyAspAArgLysIleuProIleuAlaGIyIleuValIleuLeuYrSer 76
 QY 184 GGCTACGCGCTTCGTGAGACTACCCCGAGCAGAACTGGGGCCATCCGCGGCATCGAGACCTTC 243
 Db 77 GIyTYrAlaIlePheValAspTYrProAspGIuAsnThrAlaIleAArgAlaIleGIuThrLeu 96
 QY 244 TCGGGTAAAGTGGAAATTGCATGGGAAATATATGGAAGTTGATTATCTACAGTCTCTAAAAG 303
 Db 97 SerGIyLysValGIuIleuAsnIleGIyLysIleMetGIuValAspTYrSerValSerLys 116
 QY 304 CTAAAGGAGAGGAAATTCAGATTTCGAAACATCCCTCAGCTGCAGTGGAGGTGTG 363
 Db 117 LeuAArgSerAArgLysIleGIuIleAArgAsnIleProProIleuGIuIleAsnThrIleu 136
 QY 364 GATGACTTTGGCTCAATATATGGAGACAGTGGAAATGTGGAAACAGTCAACAGACACA 423
 Db 137 AspGIyIleuLeuAlaGIuIleYrGIyThrValGIuAsnValGIuIleValAsnThrAspThr 156
 QY 424 GAAACCGCCCTGTCAACGTACATATGACAAAGAGAAAGCAAAATATGCCATGGAG 483
 Db 157 GIuThrAlaValIleAsnValThrTYrAlaThrAArgGIuIleAlaLysIleAlaMetGIu 176
 QY 484 AAGCTTAAGCGGCACTAGTTTGAAGACTACTCTTCAAGATTTCTCATCCCGAGTAA 543
 Db 177 IyIleuSerGIyIleGIuIleuAsnTYrSerPheLysIleSerTYrIleProAspGIu 196

QY	544	GAGGTGAGCTCCCTTGCGCCCTCAGCGAGCCAGCTGTGGGACCACTCTTCCCGGGAG	603
Db	197	GIuValSerSerProSerProProGlnArgAlaGlnArgGlyAspHisSerSerArgGlu	216
QY	604	CAAGGCCACGCCCTGGGGGCACTTCACAGGCCAGACAGATTGAATTCGCCGTGGCATC	663
Db	217	GInGlyHisAlaArgProGlyGlyThrSerGlnAlaArgGlnIleAspPheProLeuAaGlyIle	236
QY	664	CTGGTCCCAACCAAGTTGTGTGGCATCATCGGAAAGAGGGCTTGACCATTAAGAAC	723
Db	237	LeuValProThrGlnPheValGlyAlaIleIleGlySerGlnGlyLeuThrIleLysAsn	256
QY	724	ATCATTAAGCAGACCCAGTCCCGGTGTGATATTCATATGAAAAAGAACTGTGGAGCTGCA	783
Db	257	IleThrLysGlnThrGlnSerArgValAspIleHisArgLysGlnAsnSerGlyAlaAla	276
QY	784	GAGAAAGCCTGCACCATTCATCCAGCCCAAGGGGGACTTCTGAAGCATCCGATGATT	843
Db	277	GIuLysProValThrIleHisAlaThrProGlnGlyThrSerGlyAlaCysArgMetIle	296
QY	844	CTTGAAATCATGCGAAAAAGAGCGAGATGAGACCAACTAGCCGGAAGATTCCTCTGAA	903
Db	297	LeuGlnIleMetGlnLysGlnAlaAspGlnThrLysLeuAlaGlnGluIleProLeuLys	316
QY	904	ATCTTGGCACAACAATGGCTTGTTGGTGAAGAATGATTTGGAAAAAGAGCGAGAAATTTGAAG	963
Db	317	IleLeuAlaHisAlaGlnGlyLeuValGlyArgLeuIleGlyLysGlnGlyArgAlaSerLeuLys	336
QY	964	AAAAATTGAACATGAAACAGGGACCAAGATTAACAATCTCATCTTTGCGAGATTGAGCAT	1023
Db	337	LysIleGlnHisSGlnThrGlyThrLysIleThrIleSerSerLeuGlnAspLeuSerIle	356
QY	1024	TACAAACCCGGAAGAAACCATCATCACTGTGAAGGGCACAGTTGAGGCTGTGCACAGTGTGAG	1083
Db	357	TyrAsnProGlnArgThrIleThrValLysGlyThrValGlnAlaCysAlaSerHisGln	376
QY	1084	ATAGAGATTATGAAGAAGCTCGTAGGCTTTGAAAATGATATGCTGGCTGTTAAACAAC	1143
Db	377	IleGlnIleMetLysLysLeuAlaArgGlnAlaPheGlnAsnAspMetIleuAlaAsnThr	396
QY	1144	CACCTCCGATACTTCTCCAGCCTGATACCCCATCAACCAAGTTGGCCGTTGCCGCATCAT	1203
Db	397	HisSerGlyTyrPheSerSerLeuIleTyrProHisAlaGlnPheGlyProPheProHis	416
QY	1204	CACCTCTATTCAGAGCAGAGAGATTGTGATCTCTTCATCCCAACCCAGCTGTGGCGCC	1263
Db	417	HisSerTyrProGlnGlnGlnIleValAsnLeuPheIleProThrGlnAlaValGlyAla	436
QY	1264	ATCATCGGGAAGAGGGGGCACACATCAAAACAGCTGGCGAGATTGGCGGAGCCTCTATC	1322
Db	437	IleIleGlyLysLysGlyAlaHisIleLysGlnLeuAlaArgPheAlaGlyAlaSerIle	456
QY	1324	AAGATTGGCCCCCGGGAAGAGGCCACAGCTCAGCGAAAGATGTCATCATCAACGGGCA	1383
Db	457	LysIleAlaProAlaGlnGlyProAspAlaSerGluArgMetValIleIleThrGlyPro	476
QY	1384	CCGGAAGCCCACTTCAAGGCCCAAGGACGATCTTTGGGAAATCTGAAGAAGAAAACTTC	1443
Db	477	ProGlnAlaGlnPheLysAlaGlnGlyArgIlePheGlyLysLeuLysGlnGlnAsnPhe	496
QY	1444	TTTAATCCCAAGAAAGATGTAAGCTGGAAAGCGATATCAAGAGTCCCTCTCCACAGCT	1502
Db	497	PheAsnProLysGlnGlnValLysLeuGlnAlaHisIleArgValProSerSerThrAla	516
QY	1504	GGCCGGGTGATTGGCAAGAGTGGCAAGACCTGGAACGATCTGACAACTTAACAGTGA	1563
Db	517	GlyArgValIleGlyLysGlyLysLysThrValAsnGlnLeuGlnAsnLeuThrSerAla	536
QY	1564	GAAATCATGTGCTGTGTGACCAAAACCCAGATGAAAAAGAGAAATGATTCGTCAGAAT	1622
Db	537	GIuValIleValProArgAspGlnThrProAspGlnAsnGlnGlnValIleValArgIle	556

QY	1624	ATCGGCGCACTTCTTTGCTTAGCCAGACTGCACGCGCGAGATAGAGGAAATTGTACAAACG	1683
Db	557	IIeGlyhAspheePhAlaseRgInthmrAlaGlnArgLysIIeArgGluIleValGInGln	576
QY	1684	GTGAAGCAGCAGAGACAGAAATACCTCAGGGAGTGCCTCAGCAGCGACGAAG	1737
Db	577	ValLybGInGInGInGInLmLyStyrProGInGlyValAlaSerGInArgSerIys	594
RESULT 2			
ID	ABG96346	standard; protein; 556 AA.	
AC	ABG96346;		
XX			
DT	11-DEC-2002	(first entry)	
XX			
DE	Human ovarian cancer marker M452.		
XX			
KW	Human; ovarian cancer; marker; cancer; familial history; brain disorder;		
KW	central nervous system disorder; bacterial meningitis; viral meningitis;		
KW	Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus;		
KW	brain herniation; inflammation; encephalitis; testicular disorder;		
KW	nontuberculous granulomatous orchitis; connective tissue disorder;		
KW	heart disorder; ischaemic heart disease; atherosclerosis; neoplasm;		
KW	histological type; carcinogenic; ovarian cancer marker.		
OS	Homo sapiens.		
XX			
PN	WO200271928-A2.		
XX			
PD	19-SEP-2002.		
XX			
PF	14-MAR-2002; 2002WO-US007826.		
XX			
PR	14-MAR-2001; 2001US-0276025P.		
PR	14-MAR-2001; 2001US-0276026P.		
PR	10-AUG-2001; 2001US-0311732P.		
PR	19-SEP-2001; 2001US-0323580P.		
PR	26-SEP-2001; 2001US-0324967P.		
PR	26-SEP-2001; 2001US-0325102P.		
PR	26-SEP-2001; 2001US-0325149P.		
XX			
PA	(MILL-) MILLENNIUM PHARM INC.		
XX			
PI	Morahan JE, Gamaavarapu M, Hoersch S, Kamatkar S, Kovatis SG;		
PI	Meyers RE, Morrisey MP, Olandt RJ, Sen A, Viehey PO, Mills GB;		
PI	Bast RC, Lu K, Schmandt RE, Zhao X, Glatz K;		
XX			
DR	WPI; 2002-723277/78.		
XX	N-PSDB; ABST6442.		
PT			
PT	Assessing whether a patient is afflicted with ovarian cancer, useful in		
PT	assessing the stage or progression of the disease, comprises comparing		
PT	the expression level of a cancer marker in a sample from a patient and		
PT	from a non cancer patient.		
XX			
PS	Disclosure; Page 263-264; 481pp; English.		
XX			
CC	The present invention relates to a new method for assessing whether a		
CC	patient is afflicted with ovarian cancer. The method involves comparing		
CC	the expression level of a marker in a patient sample and the normal level		
CC	of expression of the marker in a control non-ovarian cancer sample, where		
CC	the marker is selected from 363 cancer markers described in the		
CC	specification. The method of the invention is useful in diagnosing or		
CC	characterising cancer, in detecting the presence of cancer as early as		
CC	possible, and the recurrence of ovarian cancer. The method may also be of		
CC	particular use with patients having an enhanced risk of developing		
CC	ovarian cancer (e.g. patients having a familial history of ovarian		
CC	cancer). The cancer markers may be used in the management and treatment		
CC	of e.g. brain and central nervous system disorders (e.g. bacterial and		
CC	viral meningitis, Alzheimer's disease or Parkinson's disease), brain		
CC	disorders (e.g. cerebral oedema, hydrocephalus or brain herniations),		
CC	inflammations (e.g. bacterial or viral meningitis or encephalitis),		

CC	testicular disorders (e.g. non-tuberculous granulomatous orchitis),	
CC	connective tissue disorders), or heart disorders (e.g. ischemic heart	
CC	disease or atherosclerosis). The compositions and methods may also be	
CC	used in assessing the histological type of neoplasm associated with	
CC	ovarian cancer, monitoring the progression of ovarian cancer, determining	
CC	whether ovarian cancer has metastasized or is likely to metastasize,	
CC	selecting a composition for inhibiting ovarian cancer, assessing the	
CC	ovarian carcinogenic potential of a compound, or inhibiting ovarian	
CC	cancer or at risk of developing ovarian cancer. The present amino acid	
CC	sequence represents one of the ovarian cancer markers described in the	
XX	invention.	
XX		
SQ	Sequence 556 AA:	
Alignment Scores:		
Pred. No.:	1,03e-254	
Score:	2838.00	
Percent Similarity:	100.00%	
Best Local Similarity:	100.00%	
Query Match:	48.22%	
DB:	5 Gaps: 0	
US-09-270-437D-8 (1-3283) x ABG96346 (1-556)		
QY	70 ATGATGAACAAGCTTTTACATCGGGAAACTGAAGCCCCCGGTCAACGCCGACGACTTCGGG	Length: 556
Dd	1 MetMetAsnIlyLeuThrIleGlyIysnLeuSerProAlaValThrlAlaSpAspLeuArg	Matches: 556
QY	130 CAGCTCTTTGGGGACAGGAAGTGCCCTCGGGGGACAGGTCTCTGTAAGTCCGGCTAC	Mismatches: 0
Dd	21 GlmLeuPheGlyAspArgIyLeuProLeuAlaGlyIValLeuLeuIySerGlyTyr	Indels: 0
QY	190 GCCTTCGAGGACTACCCCGACAGAACTGGGGCATCCGGCCAATGAGACCCTCTGGGT	Gaps: 0
Dd	41 AlaPheValAspTyrProaspGlnsmTrpAlalleArgAlalleGluThrIleuSerGly	
QY	250 AAGTGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAG	
Dd	61 LysValIGluLeuHieGlyLysIleMetGIuValAspTyrSerValSerIySylLeuArg	
QY	310 AGCAGGAAATTCAGATTGAAAACATCCCTCTCACCTGCAGTGGAGGGTGTGGATGA	
Dd	81 SerArgIySileGlnlleArgAsnillePropohIsleuGlnlntPrpluValleuAspGly	
QY	370 CTTTGGGTCATATGGGACAGTGGAGMATGGGAAACAAGTCAACACACAGAAACC	
Dd	101 LeuleuAlaGlnTyrglyThrValGluSenValIGluGlnValAsnThrAspThrGluThr	
QY	430 GCCGTGTCAACGTCAACATATGCAACAAGAAGAAAGCAAATAATGSCATGGAGAAGCTA	
Dd	121 AlaValValAsnValThrTyrAlahThrArgIuGluAlalysIleAlaMetGIuIySleu	
QY	490 AGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTCTCTACATCCCGGATGAAGAGTG	
Dd	141 SerGIyHsgInPheGluAsnTySerPheIySileSerTyrlleProaspGluGluVal	
QY	550 AGCTCCCTTGCGCCCCCTCAGAGAGCCAGCGTGGGACACATCTTCCCAGACAGGC	
Dd	161 SerSerProSerProProGlnArgAlaGlnIlaArgGlyAspHisSerSerArgIuGlnGly	
QY	610 CACGGCCCTGGGGGACACTTTCAGAGCCAGAACAGATTGATTCCCGGTGGAGATCCGCGTC	
Dd	181 HisAlaProGlyGlyThrSerGlnAlaArgGlnlleAspPheProLeuArgIleLeuVal	
QY	670 CCCACCCAGTTTGTGTGCCATCATCGGAAAGAGGGCTTGACCATAAAGAACATCACT	
Dd	201 ProThrGlnPheValGlyAlaIlelleGlyLysGluGlyLeuThrIleIySsnlleThr	
QY	730 AAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAAAGAACTCTGGAGCTCGAGAGAAG	
Dd	221 LysGlnThrGlnSerArgValAspIleHisArgIySgluAsnSerGlyYAlaIlaGluIyS	
QY	790 CTGTGACATCATGATGCACCCCAAGAGGGGACTTCTGAAGCATGCGCATGATTCTTGA	

```

Db      241 ProValThrIleHisAlaThrProGluGlyThrSerGluAlaCysArgMetIleLeuGlu 260
QY      850 ATCATGCAAGAAAGAGCGATGAGACCAACTAGCCGAGAGATTCTCTGAAAATCTTG 909
Db      261 IleMetGlnysGluAlaAspGluThrIleSerLeuAlaGluGluIleProLeuIleLeu 280
QY      910 GCACAGAAATGGCTTGGTGGAAAGCTGATGGAAAAGAGAGCAAAATTTAGAGAAATT 969
Db      281 AlaHisasnGlyLeuValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLysIle 300
QY      970 GAACATGAAACAGGGACCCAGATACAAATCATCTTTCAGAGATTGAGCATATACAC 1029
Db      301 GlnHsGluThrGlyThrIleIleThrIleSerSerLeuGlnAspLeuSerIleTryasn 320
QY      1030 CCGGAAAGAACCATCATCTGTGAGGGCAAGTTGAGGCTGTGCCAGTGTGAGATAG 1089
Db      321 ProGluArgThrIleThrValIleGlyThrValGluAlaCysAlaSerAlaGluIleGlu 340
QY      1090 ATTATGAGAGAGCTGCGTAGGCGCTTGAATAATGATATGCTGGCTGTTAAACCCACTCC 1149
Db      341 IleMetLysLysLeuArgGluAlaPheGluAsnMetLeuAlaValAsnThrIleSer 360
QY      1150 GGATACCTTCACAGCCTGTACCCCATCATCACAGTTTGCCGCTTCCGCATCATCACTCT 1209
Db      361 GlyTyrPheSerSerLeuTyrProHisIleGlnPheGlyProPheProHisIleHisSer 380
QY      1210 TATCCAGAGCGAGAGATTGTGATCTCTTCATCCCAACCCAGGCTGTGGCGCCATCATC 1269
Db      381 TyrProGluGlnGluIleValAsnLeuPheIleProThrGlnAlaValGlyAlaIleIle 400
QY      1270 GGGAGAGAGGGGGGACACATCAACAGCGCGGAGATTCCGCGGAGCCTCTCAAGATT 1329
Db      401 GlyLysLysGlyAlaHisIleLysGlnLeuAlaAspPheAlaGlyAlaSerIleLysIle 420
QY      1330 GCCCTGCGGAGAGGCCCAAGCCTGACCGAAGGATGGTCATCATCACCGGCGCCACCGGAA 1389
Db      421 AlaProAlaGluGlyProAspValSerGluArgMetValIleIleThrGlyProProGlu 440
QY      1390 GCCCAGTTCAAGGCCCGAGGAGCGATCTTTGGGAAACTGAAAGAGAGAAATCTTTTAAAC 1449
Db      441 AlaGlnPheLysAlaGlnIleArgIlePheGlyLysLeuLysGluGluAsnLeuPheAsn 460
QY      1450 CCCAAGAGAGAGTAGAGCTGAGAGCGCATATCAGAGTCCCTCTCCACAGCTGGCCGAG 1509
Db      461 ProLysGluGlnValLysLeuGluAlaHisIleArgValProSerSerThrAlaGlyArg 480
QY      1510 GTGATTGGCAAGGTTGGCAAGACCGTGAACGAACTGACAGAACTTAAACAGTGCAGAAATC 1569
Db      481 ValIleIleGlyLysGlyLysThrValAsnGluLeuGlnAsnLeuThrSerAlaGluVal 500
QY      1570 ATCGTGCCTCGTAGCAAAACGCCAAGTGAATAATGAGAGAGTATCGTCGATTTATCGGG 1629
Db      501 IleValProArgAspGlnThrProAspGluAsnGluGluValIleValArgIleIleGly 520
QY      1630 CACTTCTTTGCTAGACGACAGCTGACAGCGCAGAGATCAGAGAAATTTAACAACAGTGAAG 1689
Db      521 HisPhePheAlaSerGlnThrAlaGlnIleArgLysIleArgGluIleValGlnGlnValLys 540
QY      1690 CAGCAGAGACAGAAATACCTCTCAGGGAGTGCCTCTCAGCGCAGCAAG 1737
Db      541 GlnGlnGlnGlnLysTyrProGlnGlyValAlaSerGlnArgSerLys 556

```

```

KW      Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX      food supplement; medical imaging; diagnostic; genetic disorder.
OS      Homo sapiens.
PN      WO200175067-A2.
PD      11-OCT-2001.
PE      30-MAR-2001; 2001WO-US008631.
PR      31-MAR-2000; 2000US-00540217.
PR      23-AUG-2000; 2000US-00649167.
PA      (HYSE-) HYSEQ INC.
PI      Dmanac RT, Liu C, Tang YT;
XX      WPI; 2001-639362/73.
DR      N-PSDB; AAS70981.
PT      New isolated polynucleotide and encoded polypeptides, useful in
PT      diagnostics, forensics, gene mapping, identification of mutations
PT      responsible for genetic disorders or other traits and to assess
PT      biodiversity.
PS      Claim 20; SEQ ID NO 37153; 103bp; English.
XX      The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC      sequences. (I) is useful as hybridisation probes, polymerase chain
CC      reaction (PCR) primers, oligomers, and for chromosome and gene mapping;
CC      and in recombinant production of (II). The polynucleotides are also used
CC      in diagnostics as expressed sequence tags for identifying expressed
CC      genes. (I) is useful in gene therapy techniques to restore normal
CC      activity of (II) or to treat disease states involving (II). (II) is
CC      useful for generating antibodies against it; detecting or quantitating a
CC      polypeptide in tissue, as molecular weight markers and as a food
CC      supplement. (II) and its binding partners are useful in medical imaging
CC      of sites expressing (II). (I) and (II) are useful for treating disorders
CC      involving aberrant protein expression or biological activity. The
CC      polypeptide and polynucleotide sequences have applications in
CC      diagnostics, forensics, gene mapping, identification of mutations
CC      responsible for genetic disorders or other traits to assess biodiversity
CC      and to produce other types of data and products dependent on DNA and
CC      amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC      amino acid sequences of the invention. Note: The sequence data for this
CC      patent did not appear in the printed specification, but was obtained in
CC      electronic format directly from WIPO at
CC      ftp.wipo.int/pub/published_pct_sequences
XX      SQ      Sequence 614 AA;

```

Alignment Scores:

Pred. No.:	1-5e-254	Length:	614
Score:	2836.50	Matches:	567
Percent Similarity:	92.05%	Conservative:	0
Best Local Similarity:	92.05%	Mismatches:	4
Query Match:	48.19%	Indels:	45
DB:	4	Gaps:	1

```

US-09-270-437D-8 (1-3283) x ABG06794 (1-614)
QY      23 GCCGGGTACCGGGCCCGGGAGACCCGCGGCTCTCGGGGAGAGACGAGATGACAAAC 82
Db      1 AlaGlyTyrArgArg-ArgGluProArgPro-GlnIlyAsArgArgMetMetAsnLysI 20
QY      83 TTTCATCGGGAACCTGAGACCCCGCGTCAACCGGCGAGACCTCGGGGAGCTCTTGGGG 142
Db      20 eurlYrIleGlyAsnLeuSerProAlaValThrAlaAspAspLeuArgGlnLeuPheIYA 40
QY      143 ACAGAGAGCTGCCCTGCGGGGACAGATCTCTGAACTCGGCTACGCTTGTGAGACT 202
Db      40 sPArgLysLeuProLeuAlaGlyGlnValLeuLeuLysSerGlyTyrAlaPheValAspI 60

```

RESULT 3
ABG06794 standard; protein; 614 AA.
AC ABG06794;
XX 13-FEB-2002 (first entry)
DT Novel human diagnostic protein #6785.
XX

QY 203 ACCCGACCAAGACTGGGCGCATCCGCCCATCGAGACCCCTCTCGGGTAAAGTGAATTGC 262
Db 60 YRPRGASPGlnsntPrpAlaIleArgAlaIleGluThrLeuSerGlyysValGluLeuH 80
QY 263 ATGGGAAATCATGGAAGTTGATTACTGAGTCTTAAAAAGCTAAAGAGCAGGAAATTC 322
Db 80 ISGlyLysIleMetGluValAspTyrSerValSerLysLeuArgSerArgLysIleG 100
QY 323 AGATTGGAACATCCCTCTCACCCTGCGAGGTGGAGTGGATGGACTTTGGTGCAT 382
Db 100 InIleArgAsnIleProProHsiLeuGlnTyrGluValLeuAspGlyLeuLeuIleGlnT 120
QY 383 ATGGACAGTGGAGAAATGTGGAACAAGTCAACACAGACAGAGAAACCCCGTGTCAAG 442
Db 120 YRGLYThrValGluAsnValGluGlnValAsnThrAspThrGluThrAlaValAlaAsnV 140
QY 443 TCACATATGCAACAGAGAGAGAGCAAAATAGCCATGAGAGAGCTAACGGCGCATCAGT 502
Db 140 alThrTyrAlaThrArgGluGluIuAlaLysIleAlaMetGluLysLeuSerGlyHsiGlnP 160
QY 503 TTGGAAGTACTCCCTTCAAGATTTCCTACATCCCGGATGAAAGAGGTGAGCTCCCTTGC 562
Db 160 heGluAsnTyrSerPheLysIleSerTyrIleProAspGluGluValSerSerProSerP 180
QY 563 CCCCTCAGCGAGCCGAGCGTGGGAGCACTCTCCCGGAGCAAGCGCCCTGCGG 622
Db 180 roProGlnArgAlaGlnArgGlyAspHsiSerSerArgGluGlnGlnHsiAlaProGlyG 200
QY 623 GCACCTTTCAGGCGCAACAGATTGATTTCCCGTGGAGATCCGTGCCCCACCGATTGG 682
Db 200 LyrThrSerGlnAlaArgGlnIleAspPheProLeuArgIleLeuValProThrGlnPheV 220
QY 683 TTGGTGCATCAGTGGGAAAGAGGGCTTGACCCATAAAGAACATCACTAAGACGCCACT 742
Db 220 alGlyAlaIleIleGlyLysGlnGlyLeuThrIleLysAsnIleThrLysGlnThrGlnS 240
QY 743 CCCGGTAGATATCATAGAAAAGAGAACTGTGAGGCTGCAGAGAGCGCTGCACATCC 802
Db 240 erArgValaPrIleHsiArgLysGluAsnSerGlyAlaAlaGlnLysProAlaThrIleH 260
QY 803 ATGCCACCCAGAGGGGAGCTTGTGAGCATGCCGATGATTTCTTGAATCATGAGGAAG 862
Db 260 ISAlaThrProGluGlyThrSerGluAlaCysArgMetIleLeuGlnIleMetGlnLysG 280
QY 863 AGGACGATGAGCAACAACTGCGGAGAGAGATTCCTCTGAAAAATTTGGCACACATGGCT 922
Db 280 IuAlaAspGluThrLysLeuAlaGlnIleProLeuLysIleLeuAlaHsiAsnGlyL 300
QY 923 TGGTTGGAGACTGATTGAAAAGAGGAGCAAAATTTGAAGAAATGAACTGAACAG 982
Db 300 euValGlyArgLeuIleGlyLysGlnGlyLysArgAsnLeuLysLysIleGlnHsiGlnThrG 320
QY 983 GGAACCAAGATPAACAATTCATCTTTTGAGAGATTGACATATACAAACCCGAGAAAGACCA 1042
Db 320 LyrThrLysIleThrIleSerSerLeuGlnAspLeuSerIleTyrAsnProGluArgHrI 340
QY 1043 TCACCTGGAGGGGACAGCTTGAGGCCCTGTGCCAGTGTGGATGATAGATATTAAGAAGC 1102
Db 340 IeThrValLysGlyThrValGluAlaCysAlaSerAlaGlnIleGlnIleMetLysIyl 360
QY 1103 TGGGTAGAGGCTTTGAAAATGATATGCTGCTGTAAAC----- 1140
Db 360 euArgGluAlaPrGlnAsnAspMetLeuAlaValAsnGlnGlnAlaAsnLeuIleProG 380
QY 1140 ----- 1140
Db 380 LyrLeuAsnLeuSerAlaLeuGlyIlePheSerThrGlyLeuSerValLeuSerProProA 400
QY 1141 -----ACCACTCCGAGAT 1153
Db 400 IeGlyProArgGlyAlaProProAlaAlaProTyrHsiProPheThrThrHsiSerGlyT 420

QY 1154 ACTTCTCAGGCTGTATCCCCCATCAGCAGTTTGGCCCGTTCGCCGATCATCTTATTC 1213
Db 420 YRPRSerSerLeuTyrProHsiHsiGlnPheGlyProPheProHsiHsiSerTyrP 440
QY 1214 CAGACAGAGATTGTGAATCTCTTCATCCCAACCCAGGCTGTGGGCGCCATCATCGGGA 1273
Db 440 roGluGlnGlnIleValAsnLeuPheIleProThrGlnAlaValGlyAlaIleIleGlyL 460
QY 1274 AGAAGGGGGCACACATCAACAGCTGGCGAGATTGGCCGAGCCTTATCAAGATTGCC 1333
Db 460 YslYsgIylAlaHsiIleLysGlnLeuAlaArgPheAlaGlyAlaSerIleLysIleAlaP 480
QY 1334 CTGGGGAAGGCCAGAGGTCAGCGGAAGAGATGTCATATACCGGGGCCACCGGAAGCCC 1393
Db 480 roAlaGluGlyProAspValSerGluArgMetValIleIleThrGlyProProGluAlaG 500
QY 1394 AGTTCAAGGCCACAGAGCGGATCTTTGGGAAACTGAAAGAGAAAACTTTTAAACCCCA 1453
Db 500 InPheLysAlaGlnGlyArgIlePheGlyLysLeuLysGluGluAsnPheAsnProL 520
QY 1454 AAGAAAGATGAGACCTGGAAGCGCATATCAGAGTGCCCTTTCACAGCTGGCCGGTGA 1513
Db 520 YsgIuGluValLysLeuGluAlaHsiIleArgValProSerSerThrAlaGlyArgValI 540
QY 1514 TTGGCAAAAGGTGGCAAGACCGTGAACGACCTGCAGAACTTAACAGTGCAGAACTCATCG 1573
Db 540 IeGlyLysGlyGlyLysThrValAsnGlnLeuGlnAsnLeuThrSerAlaGluValIleV 560
QY 1574 TGCCCTGTGACCAACCCAGATGAAATGAGAAATGATCGTGAATTAATCGGCGACT 1633
Db 560 alProArgAspGlnThrProAspGluAsnGlnGluValIleValArgIleIleGlyHsiP 580
QY 1634 TCTTTGCTAGCCAGACTGTCACAGCGCAAGATCAGGGAATTTGACACAGGTGAAGAGC 1693
Db 580 hePheAlaSerGlnThrAlaGlnArgLysIleArgGlnIleValGlnGlnValLysGlnG 600
QY 1694 AGGACGAAATACCTCCAGGAGTCCGCTCAGACGCCACGACAG 1737
Db 600 InGluGlnLysTyrProGlnGlyValAlaSerGlnArgSerLys 614

RESULT 4
AAU16163
ID AAU16163 standard; protein; 620 AA.
XX AC
XX AAU16163;
XX 07-NOV-2001 (first entry)
XX DT
XX DE
XX Human novel secreted protein, Seq ID 1116.
KW Human; immunosuppressive; antiarthritic; antirheumatic; cyostatic;
KW cardiatic; vasotrophic; cerebroprotective; nootropic; neuroprotective;
KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
OS Homo sapiens.
XX XX
XX EN WO200155322-A2.
XX PD
XX 02-AUG-2001.
XX XX
XX 17-JAN-2001; 2001WO-US001341.
XX PF
XX 31-JAN-2000; 2000US-0179065P.
XX PR 04-FEB-2000; 2000US-0180628P.
XX PR 24-FEB-2000; 2000US-0184664P.
XX PR 02-MAR-2000; 2000US-0186350P.
XX PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216889P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226272P.
PR 22-AUG-2000; 2000US-0226881P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236127P.
PR 29-SEP-2000; 2000US-023617P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.

PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0251989P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251899P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-488783/53.
XX N-PFSD; AAS26150.
XX
XX New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX
XX Claim 11, SEQ ID NO 1116; 980pp; English.
XX
XX The invention relates to isolated nucleic acid molecules and their

CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
 CC in diagnosing a pathological condition or susceptibility to a
 CC pathological condition. Antibodies to the proteins can also be used in
 CC alleviating symptoms associated with the disorders and in diagnostic
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
 CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angioneu-
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
 CC and many other disorders listed in the specification. The polypeptides
 CC can also be used to aid wound healing and epithelial cell proliferation,
 CC to prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence represents a novel secreted protein of the invention. Note: The
 CC sequence data for this patent did not form part of the printed

Alignment Scores:

Pred. No.:	6,82e-252	Length:	620
Score:	2808.00	Matches:	560
Percent Similarity:	90.78%	Conservative:	1
Best Local Similarity:	90.61%	Mismatches:	11
Query Match:	47.71%	Indels:	46
DB:	4	Gaps:	2

US-09-270-437d-8 (1-3283) x ANU16163 (1-620)

```

QY 22 CGCGGCGTACCGGCGCGGCGGCGCTTCGCGGGAAGACGG-----ATG 72
DB 3 ArgArgTyrAlaCysArgTyrArgSerGlyLeuProGlySerThrHisAlaSerGlyMet 22
QY 73 ATGACCAAGCTTTACATCGGGAACCTGAGCCCGCGCTCACCGCGACGACTCCGGCAG 132
DB 23 MetGlnLysLeuTyrLeuGlyAsnLeuSerProAlaValThrAlaAspLeuArgGln 42
QY 133 CTCTTTGGGGAACAGAAAGCTGCCCCCTGGCGGACAGATCTCGTGAAGTCCGGCTACG 192
DB 43 LeuPheGlyAspArgGlyLeuProLeuAlaGlyGlnValLeuLeuLysSerGlyTyrAla 62
QY 193 TTGGTGAATACCCCGGACCGAAATTGGGCCAATCCGCCCATCGAGACCTTCGGGTAA 252
DB 63 PheValAspTyrProAspGlnAsnTyrAlaIleArgAlaIleGluThrLeuSerGlyLys 82
QY 253 GTGGAATTGCATGGGAAATCATGGAAGTTGATTACTCAGTCTTAAAGCTTAAGAGC 312
DB 83 ValGluLeuHisGlyLysIleMetGlnValAspTyrSerValSerLysLysLeuArgSer 102
QY 313 AGGAAATTCAGATTCCGAACATCCCTCCTCACCCTGAGTGGAGGTGGTGAAGACT 372
DB 103 ArgLysIleGlnIleArgAsnIleProProHisLeuGlnThrGluValLeuAspGlyLeu 122
QY 373 TTGGCTCAATATGGGACAGTGGAGATGTGGAAACAATCAACACAGACAGAAACCGCC 432
DB 123 LeuAlaGlnTyrIleGlyThrValGluAsnValGluGlnValAsnThrAspThrGluThrAla 142
QY 433 GTTGTCAACGTCACATATGCAACAGAGAAAGCAAAATTAAGCTGGAAGGCTTAAGC 492
DB 143 ValValAsnValThrTyrAlaThrArgGluGlnAlaLysIleAlaMetGlyLysLeuSer 162
QY 493 GGGCATCAGTTTGAGAACTACTCCTTCAGATTTCCTACATCCCGGATGAAGAGTGAGC 552
DB 163 GlyHisGlnPheGlyAsnTyrSerPheLysIleSerTyrIleProAspGluGlnValSer 182
QY 553 TCCGCTTTCGCCCCCTCAGCGAGCCACAGCGTGGGAGACCACTTTCGCGGAGCAAGCCAC 612
DB 183 SerProSerProProGlnArgAlaGlnArgGlyAspHisSerSerArgGluGlnGlyHis 202

```

```

QY 613 GCCCTGGGGGACATCTTTCAGGCCAGACAGATTGATTTCCGCTGCGGATCCTGTGCCCC 672
DB 203 AlaProGlyGlyThrSerGlnAlaArgGlnIleAspPheProLeuArgIleLeuValPro 222
QY 673 ACCGAGTTGTGGGCGCATCATCGGAAAGAGGGCTTGACCATTAAGAACTACTAG 732
DB 223 ThrGlnPheValGlyAlaIleIleGlyLysGluGlyLeuThrIleLysAsnIleThrLys 242
QY 733 CAGACCCAGTCCCGGCTGATATCATATGAAAGAAAGAACTCGAGCTGCAGAGAACCT 792
DB 243 GlnThrGlnSerArgValAspIleHisArgLysGluAsnSerGlyAlaAlaGluLysPro 262
QY 793 GTACACATCATGCCACCCCGAGGCGGACTTCTGAAGCATGCCGATGATCTTGAATC 852
DB 263 ValThrIleHisAlaThrProGluGlyThrSerGluAlaCysArgMetIleLeuGluIle 282
QY 853 ATGACGAAAGGGGACAGATGAGAACAAATACCCGAAAGATTCCTCTGAAATCTTGCGCA 912
DB 283 MetGlnLysGluAlaAspGluThrLysLeuAlaGluGluIleProLeuLysIleLeuAla 302
QY 913 CACATGCGCTGTGGAAAGACTGATTGGAAGAGAGGACAGAAATTGAAGAAATTTGAA 972
DB 303 HisAsnGlyLeuValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLysLysIleGlu 322
QY 973 CATGAACAGGGACCGAAGATTAACAATCTCATCTTTGACAGATTGAGCATATACACCG 1032
DB 323 HisGluThrGlyThrLysIleThrLysSerSerLeuGlnAspLeuSerIleTyrAsnPro 342
QY 1033 GAAAGAACCATCATCTGGAAGGGGACAGCTTGAGGCCCTGTGCATGCTGAGATAGAGATT 1092
DB 343 GluArgThrIleThrValLysGlyThrValGlnAlaCysAlaSerAlaGluIleGluIle 362
QY 1093 ATGAAAGAGCTGCGTGAAGGCTTTGAAATGATATGCGGCTGTTAAC----- 1140
DB 363 MetLysLysLeuArgGluAlaPheGluAsnAspMetLeuAlaValAsnGlnAlaAsn 382
QY 1140 ----- 1140
DB 383 LeuIleProGlyLeuAsnLeuSerAlaLeuGlyIlePheSerThrGlyLeuSerValLeu 402
QY 1141 -----ACC 1143
DB 403 SerProProAlaGlyProAspArgGlyAlaProProAlaAlaProTyrHisProPheThrThr 422
QY 1144 CACTCCGATATCTTCCAGCTGTATCCCATCAACAGATTGGCCCGCTCCGCAATCAT 1203
DB 423 HisSerGlyTyrThrPheSerSerLeuTyrProHisHisGlnPheGlyProPheProHis 442
QY 1204 CACTCTTATCCAGACAGAGATTTGAAATCTCTTATCCCAACCCAGGCTGTGGCGCC 1263
DB 443 HisSerTyrProGluGlnGluIleValAsnLeuPheIleProThrGlnAlaValGlyAla 462
QY 1264 ATCATGGGAAAGAGGGGGGACACATCAACAACTACTGCGGAGATTCGCGGAGCTTATC 1323
DB 463 IleIleGlyLysLysGlyAlaHisIleLysGlnLeuAlaArgPheAlaGlyAlaSerIle 482
QY 1324 AAGATTGCCCTCGGAGAGGCCAGACGTCAAGGAAAGATGTGCATCATCAACCGGGCCA 1383
DB 483 LysIleAlaLeuProAlaGluGlyProAspValSerGluArgMetValIleIleThrGlyPro 502
QY 1384 CCGGAAGCCCAAGTTCAAGGCCACGGAACGATCTTTGGGAAACTGAAGAGAAACTTC 1443
DB 503 ProGluAlaGlnPheLysAlaGlnGlyArgIlePheGlyLysLeuLysGluGluAsnPhe 522
QY 1444 TTTAAACCCCAAGAAAGATGAGCTGGAAGCGCATATCAAGATGCCCTTCCACAGCT 1503
DB 523 PheAsnProLysGluGluValLysLysLeuGln**HisIleArgValProSerSerThrAla 542
QY 1504 GGCCTGGGTATTTGGCAAGGTGGCAAGACGTGAAGAACTGAGAGAACTTAACAGATGCA 1563
DB 543 GlyArgValIleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuThrSerAla 582

```


Db 63 PheValAspTyrProAspGlnAsnTrpAlaIleArgAlaIleGluThrIleuSerGlyLys 82
 QY 253 GTGGAATGTCATGGGAAATCATGAGATTGATTCTCATGCTCTTAAAGCTAAGAC 312
 Db 83 ValGluLeuHisGlyLysIleMetGluValAspTyrSerValSerLysLysLeuArgSer 102
 QY 313 AGGAAATTCAGATTTCGAAACATCCCTCCTCAGCTGAGAGGTGGTGGATGGAACCT 372
 Db 103 ArgLysIleGlnIleArgAsnIleProProHisLeuGlnIleProGluValLeuAspGlyLeu 122
 QY 373 TTGGCTCAATATGGGACAGTGGAGATGTGAAACAGTCAACACAGACAGAAACCCGC 432
 Db 123 LeuAlaGlnTyrGlyThrValGluAsnValGluGlnValAsnThrAspThrGluThrAla 142
 QY 433 GTTGCAACGTCACATATGCAACAGAGAAAGAAATAGCCATGAGAGAGCTAAGC 492
 Db 143 ValValAsnValThrTyrAlaThrArgGlnGluAlaLysIleAlaMetGluLysLeuSer 162
 QY 493 GGGCATCAGTTTGAGAACTACTCTCTCAAGATTCTCATCCGGATGAAAGAGGTGAGC 552
 Db 163 GlyHisGlnPheGlnAsnTyrSerPheLysIleSerTyrIleProAspGluGluValSer 182
 QY 553 TCCCTCTGGCCCTCAGCGAGCCAGCGTGGAGCACTCTTCCGGAGCAAGGCCAC 612
 Db 183 SerProSerProGlnArgAlaGlnArgGlyAspHisSerSerArgGluGlnGlyHis 202
 QY 613 GCCCTGGGGGACATTCTCAGCGACAGACAGATTGATTTCCGGCTGGCATCTGTGCTCCC 672
 Db 203 AlaProGlyGlyIleSerGlnAlaArgGlnIleAspPheProLeuArgGlyLeuValPro 222
 QY 673 ACCGAGTTTGTGGTCATCATCGAAAGAGAGGGCTTGACCATTAAGAAATCATGACTAAG 732
 Db 223 ThrGlnPheValGlyAlaIleIleGlyLysGlnIleuThrIleLysAsnIleThrLys 242
 QY 733 CAGACCCAGTCCGGGTAGATATCATAGAAAGAGAACTCTGGAGCTGCAGAGAGCCT 792
 Db 243 GlnThrGlnSerArgValAspIleHisArgLysGluAsnSerGlyAlaAlaGluLysPro 262
 QY 793 GTACACCATGCATGCCACCCAGAGGGGACTTCTGAAGCATGCCCATGATTTCTGAATC 852
 Db 263 ValThrIleHisAlaThrProGlnGlyThrSerGlnLacysArgMetIleLeuGlnIle 282
 QY 853 ATGAGAAAGAGGCGATGAGACCAACCACTAGCCGAGAGATTCCCTGAAATCTTGCGCA 912
 Db 283 MetGlnLysGlnAlaAspGlnThrLysLeuAlaGlnIleIleProLeuLysIleLeuAla 302
 QY 913 CACAAATGCTTGTGGTGAAGACTGATTGAAAGAGAGAGCAAAATTTGAAGAAATTTGAA 972
 Db 303 HisAsnGlyLeuValGlyArgLeuIleGlyLysGlnIleArgAsnLeuLysIleGlu 322
 QY 973 CATGAACAGGGACCAAGATATCAATCTCTTTGGAGATTTGAGCATATACACCCG 1032
 Db 323 HisGlnThrGlyThrLysIleThrIleSerSerLeuGlnAspLeuSerIleTyrAsnPro 342
 QY 1033 GAAAGAACCATCACTGAGAGGCGACAGTTGAGGCGCTGTCGCCAGTGTGATAGATTAAGATT 1092
 Db 343 GluArgThrIleThrValLysGlyThrValGluAlaCysAlaSerAlaGlnIleGluIle 362
 QY 1093 ATGAAGAACTGGGTGAGGCGCTTTGAAATGATATGCTGCTGTAAAC----- 1140
 Db 363 MetLysLysLeuArgGlnAlaPheGlnAsnAspMetLeuAlaValAsnGlnGlnAlaAsn 382
 QY 1140 ----- 1140
 Db 383 LeuIleProGlyLeuAsnLeuSerAlaLeuGlyIlePheSerThrGlyLeuSerValLeu 402
 QY 1141 -----ACC 1143
 Db 403 SerProProAlaGlyProArgGlyAlaProProAlaAlaProTyrHisProPheThrThr 422
 QY 1144 CATCCGGATATCTTCTCAGCCTGTACCCCATCACCAGTTTGCCCGCTCCCGCATCAT 1203

Db 423 HisSerGlyTyrPheSerSerLeuTyrProHisHisGlnPheGlyProPheProHisHis 442
 QY 1204 CACTTATTCAGACAGAGAGATTGTAATCTTTCATGCCAACCCAGGCTGTGGGCGCC 1263
 Db 443 HisSerTyrProGlnGlnGlnIleValAsnLeuPheIleProThrGlnAlaValGlyAla 462
 QY 1264 ATCATTCGGAGAAAGAGGGGCGACACATCAAAAGCTGGGAGATTGCCGGAGCTCTATC 1323
 Db 463 IleIleGlyLysLysGlyAlaHisIleLysGlnLeuAlaArgPheAlaGlyAlaSerIle 482
 QY 1324 AAGATTGGCCCTGGCGAAAGGCCCGACAGCTCAGCCGGAAGAGGTCATCATCAACCGGCGCA 1383
 Db 483 LysIleAlaProAlaGluGlyProAspValSerGluArgMetValIleIleThrGlyPro 502
 QY 1384 CCGAAGCCCATTCAGAGCCAGGAGCGATCTTTGGGAAACTGAAGAGGAAACTTC 1443
 Db 503 ProGluAlaGlnPheLysAlaGlnGlyArgIlePheGlyLysLeuLysGluGlnAsnPhe 522
 QY 1444 TTTAACCCTCAAGAAAGAAAGTAAAGCTGGAAGCGCATATCAAGTGCCTTTCACAGCT 1503
 Db 523 PheAsnProLysGluGluValLysLeuGluHisHisIleArgValProSerSerThrAla 542
 QY 1504 GCGCGGTGATTGGCAAGGTGGAGAGCCGTGAACCGTGAACGAACTTAACGAGTGC 1563
 Db 543 GlyArgValIleGlyLysGlyLysThrValAsnGluLeuGlnAsnLeuThrSerAla 562
 QY 1564 GAAGTCATCGTCTGTGACCAACCGCAGATGAAATGAGAAATGATCGTCAAGATT 1623
 Db 563 GluValIleValProArgAspGlnThrProAspGluAsnGlnGluValIleValArgIle 582
 QY 1624 ATCGGACATCTTCTTGTAGCCAGACTGCACAGCGCAAGATCAGGAAATTTGTCACACAG 1683
 Db 583 IleGlyHisPhePheAlaSerGlnThrAlaGlnArgLysIleArgGlnIleValGlnGln 602
 QY 1684 GTGAAGCAGCAGAGCAGAAATATACCTCAGGAGATCGCTCAGCGCCAGCAAG 1737
 Db 603 ValLysGlnGlnGlnGlnLysTyrProGlnGlyAlaIleAspGlnArgSerLys 620

RESULT 6
 ID ABU89799 standard; protein: 555 AA.
 XX
 AC ABU89799;
 XX
 DT 10-JUL-2003 (first entry)
 XX
 DE Novel human protein NOV14a.
 XX
 KW Human; cytostatic; DAPK3-Agonist; DAPK3-Antagonist; cancer; NOV.
 OS Homo sapiens.
 XX
 PN WO2003031571-A2.
 XX
 PD 17-APR-2003.
 XX
 PF 02-OCT-2002; 2002MO-US031357.
 XX
 PR 05-OCT-2001; 2001US-0327454P.
 PR 09-OCT-2001; 2001US-0327917P.
 PR 09-OCT-2001; 2001US-0328029P.
 PR 09-OCT-2001; 2001US-0328056P.
 PR 12-OCT-2001; 2001US-0328849P.
 PR 15-OCT-2001; 2001US-0329414P.
 PR 17-OCT-2001; 2001US-0330142P.
 PR 22-OCT-2001; 2001US-0341058P.
 PR 24-OCT-2001; 2001US-0343629P.
 PR 29-OCT-2001; 2001US-0349575P.
 PR 01-NOV-2001; 2001US-0346357P.
 PR 25-NOV-2002; 2002US-0391342P.
 PR 01-OCT-2002; 2002US-00262445.
 XX
 PA (CURA-) CTRAGEN CORP.

XX Alsobrook JP, Burgess CE, Catterton E, Chant JS, Chaudhuri A;
PI Edinger SR, Gerlach VM, Giot U, Gorman M, Guo X, Kekuda R;
PI Mezes PS, Millet I, Ooi CE, Paturajan L, Rieger DK, Spletke RA;
PI Taupier RJ, Zehusen BD, Zhong H, Zhong M;
XX WPI; 2003-381704/36.
DR N-PSDB; ACA90176.
XX
XX New DAPK3 polypeptide, useful for preparing a composition for treating or
preventing e.g., cancer.
XX
XX Claim 2; Page 129; 253pp; English.
XX
XX The invention describes an isolated polypeptide comprising any of 33 90-
CC 1273 amino acid sequences (I) given in the specification or its mature
CC form, a sequence that is at least 95 % identical to (I), or a sequence
CC comprising one or more conservative substitutions in the amino acid
CC sequence of (I). The polypeptide is useful for preparing a composition
CC for treating or preventing e.g. cancer. This is the amino acid sequence
CC of a novel human NOV protein
CC
XX
XX Sequence 555 AA;
SQ
Alignment Scores:
Pred. No.: 4.25e-231 Length: 555
Score: 2584.50 Matches: 507
Percent Similarity: 94.60% Conservative: 19
Best Local Similarity: 91.19% Mismatches: 29
Query Match: 43.91% Indels: 1
Gaps: 1
DB:
US-09-270-437d-8 (1-3283) x AB089799 (1-555)
QY 70 ATGATGAACAACCTTACATCGGAACTGAGCCCGCGTACCGCGACGACCTCCGG 129
DB 1 MetMetAsnLysLeuPheHleGlyAsnLeuSerProAlaValThrAlaGluPheLys 20
QY 130 CAGCTCTTGGGAGACAGGAAGCTCCCTGGGAGACAGGTCTCTGTAAGTCCGGCTAC 189
DB 21 GlnLeuPheGlyAspArgLysLeuProLeuAlaGlyGlnValLeuLysSerArgTyr 40
QY 190 GCCTTCGTGGATACCCCGACGACGAGCTGGGACCTCGGACCTTCGGGT 249
DB 41 AlaPheValAspTyrProAspGlnAsnThrAlaIleArgThrIleGlnThrLeuSerGly 60
QY 250 AAAGTGAATTGCGATGGGAAATCATGGAAGTTGATTACTCACTCTTAAAGCTTAAG 309
DB 61 GlnValGlnLeuHISgilyLysIleMetGlnValAspTyrSerValSerIleLysLeuArg 80
QY 310 AGCAGGAAATTCAGATTGGAACATTCCTCTCCTCAGTGCAGTGGAGGTGTGATGA 369
DB 81 SerArgAsnIleProIleArgAsnIleProProHISLeuGlnIlePheGlnValLeuAspGly 100
QY 370 CTTTGGGCTCAATATGAGGACAGTGGAGATGTGGAACAAGTCAACAACAGACGAAAC 429
DB 101 LeuLeuAlaGlnIleTyrGlyThrValGlnAsnValGlnGlnValAsnThrArgIleThr 120
QY 430 GCCGTTGTCAACGTCATATGCAACAGAGAAAGAAAGAAATATGCGATGAGAAAGCTA 489
DB 121 AlaValValAsnValThrTyrAlaThrIleGlyGlnValIleAlaMetCysLysLeu 140
QY 490 AGCGGGCATATGATTGAAACTACTCTTCAAGATTTCACATCCCGGATGAAAGGTG 549
DB 141 SerGlyHISglnPheGlnAsnHISLysPheLysIleSerTyrIleProAspAspGlnVal 160
QY 550 AGCTCCCTTCGCGCCCTCAGCGAGCCGAGTGGGAGACCACTTTCGCGAGCAAGGC 609
DB 161 SerCysProSerProProGlnArgAlaGlnArgGlyAspHisSerSerTyrPheGlnGly 180
QY 610 CACGCGCTTGGGGGACATCTTCAGCGCCAGACAGATTGATTCCGCTGCGGATCTGGTC 669
DB 181 GlnAlaProIleGlySerSerGlnAlaArgGlnIleAspPheProLeuArgValLeuPhe 200

QY 670 CCCAGCCAGTTTGGTGGCCATCATCGAAAGAGGGCTTGACATTAAGACATCACT 729
DB 201 ProThrGlnPheValGlyAlaIleIleGlyLysGlnGlyLeuThrIleLysAsnIleThr 220
QY 730 AAGCAGAACCCAGTCCCGGATGATATTCATATGAAAAAGAACTCTGAGCTCGAGAGAG 789
DB 221 LysGlnSerArgSerArgValAspIleTyrArgGlnGlnLysSerArgAlaAlaGlnLys 240
QY 790 CCTCTCAACCATCCATGCGACCCCGAGGGGACTCTGAGAGATGCGCATGATCTTGA 849
DB 241 ProValThrMetHISAlaThrProGlnGlyThrSerGlnAlaCysArgMetIleLeuGln 260
QY 850 ATCATGAGAAAGAGGAGATGAGACCAACTAGCCGGAAGATTCCTCTGAAAACTTGT 909
DB 261 IleMetGlnLysGlnAlaAspGlnAlaLysLeuAlaGlnGlnIleProLeuLysIleLeu 280
QY 910 GCACACAATGGCTGTGTTGAAAGCTGATTTGGAAAAGAGGAGCAAAATTTGAGAAAAAT 969
DB 281 AlaHISAsnGlyLeuValGlyArgLeuIleGlyLysGlnGlyArgAsnLeuLysLysAsn 300
QY 970 GAACATGAACAGGAGACCAAGATTAACAATCTCATCTTTGACAGATTGACATATACAAC 1029
DB 301 GlnHISGlnThrGlyThrLysIleThrIleSerSerSerGlnAspLeuSerIleTyrAsn 320
QY 1030 CCGGAAAGAACCATCATCTGTGAAGGAGCAAGTTGAGCCCTGTGCAATGATAGAG 1089
DB 321 ProGlnArgThrIleThrValLysGlyThrValGlnValCysAlaSerAlaGlnIleGln 340
QY 1090 ATTATGAAGAGCTCGGAGAGCCCTTGAATAATGATATGATCGGCTGTAAACCCACTCC 1149
DB 341 IleMetLysLysLeuArgGlnAlaPheGlnAsnAspThrLeuThrValAsnThrHISpHe 360
QY 1150 GGATACCTTCGACGCTGTACCCCATGACCAAGTTTGGCCCGCTCCGATCATCACTCT 1209
DB 361 GlyTyrPheSerSerLeuTyrProHISArgGlnPheGlyProPheProHISHisSer 380
QY 1210 TATCCAGACGAGAGATTGTGAATCTTCAATCCACCCAGCTGTGGGCGGCATCAATC 1269
DB 381 TyrProGlnGlnGlnIleValAsnLeuPheIleProThrGlnGlyValGlyAlaIleIle 400
QY 1270 GGGAAAGAGGGGAGCACATCAATCAACAGCTGGGAGATTGCCGGAGCTCTTCAAGATT 1329
DB 401 GlyLysLysGlyAlaHISIleLysGlnLeuAlaAspPheValGlyAlaSerIleLysIle 420
QY 1330 GCCCGCGGAGAGCCGACGCTCAGCGAAAGAGATGTGATCATCAACCGGCGCACCGGA 1389
DB 421 AlaProAlaArgSerPro---LeuArgGlnArgLysValIleIleThrTyrProGln 439
QY 1390 GCCAGATTCAAGGCCGAGGAGCGGATCTTTGGGAAACTGAAAGAGGAAACTTCTTAAC 1449
DB 440 SerGlnPheLysAlaGlnIleArgIlePheGlyLysLeuLysGlnLysAsnPheAsn 459
QY 1450 CCGAAAGAAAGTGAAGTGAAGCGCATATCAAGAGTGCCTTCCTCCACAGCTGCCCG 1509
DB 460 ProLysGlnAspAlaLysLeuGlnThrHISIleArgValProSerSerThrAlaGlyArg 479
QY 1510 GTGATTGGCAAGAGTGGCAAGACCGTGAAGCACTGCAAGACTTAACAGTGCAGAAATC 1569
DB 480 ValIleGlyLysGlyLysThrValAsnGlnLeuGlnAsnLeuIleSerAlaGlnVal 499
QY 1570 ATGTGCTCTGTGACCAAAAGCCGAGTGAATAATGAGAAAGTATTCGTAATAATTCGG 1629
DB 500 IleValProArgAspGlnThrProAspGlnAsnGlnGlnMetIleValAlaGlyIleGly 519
QY 1630 CACTTCTTGTCTGACCAAGCTGACAGCGGCAAGATCAGGGAATTTGATCAACAGTGAAG 1689
DB 520 HisPhePheAlaSerGlnThrAlaGlnArgLysIleArgGlnIleValGlnGlnValLys 539
QY 1690 CACGAGAGCAGAAATACCTTCAGGAGTGCCTCAGGCGCAGCAAG 1737
DB 540 GlnGlnGlnGlnLysTyrProGlnGlnValAlaSerGlnArgSerLys 555

```
RESULT 7
AA30649
ID   AA30649 standard; protein; 577 AA.
XX
AC   AA30649;
XX
DT   17-NOV-1999 (first entry)
XX
DE   A murine c-myc coding region determinant binding protein.
XX
KW   c-myc coding region determinant binding protein; CRD-BP; tumor; c-myc;
KW   endonucleolytic attack; half-life; breast cancer; colon cancer;
KW   pancreatic cancer.
XX
OS   Mus musculus.
XX
PN   M09946594-A2.
XX
PD   16-SEP-1999.
XX
PE   05-MAR-1999; 99WO-US004897.
XX
PR   09-MAR-1998; 98US-0077372P.
XX
PA   (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
PI   Ross J;
XX
DR   WPI; 1999-551506/46.
DR   N-PSDB; AA210617.
XX
PT   Diagnosing presence or absence of a tumor in a human by examining c-myc
PS   coding region determinant-binding protein.
XX
PE   Example; Fig 1A-D; 79pp; English.
XX
CC   The present sequence represents a murine c-myc coding region determinant
CC   binding protein (CRD-BP). The presence or absence of a tumor can be
CC   determined by determining the levels of CRD-BP present in the suspect
CC   tissue, where the CRD-BP shields c-myc RNA from endonucleolytic attack
CC   and so prolongs its half-life. The methods are used for diagnosing
CC   presence or absence of a tumor in a human, especially breast, colon and
CC   pancreatic cancer. They are also used to inhibit cancer cell growth
XX
SQ   Sequence 577 AA:

Alignment Scores:
Pred. No.:      1.74e-170      Length:      577
Score:          1934.00        Matches:      381
Percent Similarity: 79.10%      Conservative: 77
Best Local Similarity: 65.80%      Mismatches:  95
Query Match:     32.86%        Indels:       26
DB:              2            Gaps:         5

US-09-270-437D-8 (1-3283) x AA30649 (1-577)

QY   73   ATGAACAAGGTTACATCGGGAAGTCCCGCGTCAACCGCCAGACGACTTCGGGAG 132
DB     |||||
QY   1   MetAsnysLeuYrllleGlyAsnLeuAsnGlnSerValThrProAlaAspLeuGlnLys 20
DB     |||||
QY   133  CTCTTTGGGGACGAGAGCTGCCCTGCGGAGACAGGTCCTGCTGAAGTCCGGCTACGGC 192
DB     |||||
QY   21   ValPheAlaGlnHisIleSerIyrSerGlyGlnPheLeuValIylsSerGlyTyrAla 40
DB     |||||
QY   193  TTTCGTGACTACCCCGACGACGAACTGGCCATCCGGCCATCGAGACCCCTTCGGGTAA 252
DB     |||||
QY   41   PheValAspCysProAspGlnHisIleThrAlaMetIylAlaIleGlnThrPheSerGlyLys 60
DB     |||||
QY   253  GTGGAAATTCATGAGGAAATTCATGAGTGAATTAATCTCTTAAAGCTAAAGCTAAGAGC 312
DB     |||||
QY   61   ValGlnLeuGlnIleGlySarGlnLeuGlnMetGlnHisSerValProLysIylsGlnArgSer 80
DB     |||||
QY   313  AGGAAATTCATGATTCGAAACATCCCTCTCACTGACGAGGAGGTGTTGATGAGCTT 372
```

```
DB     |||||
QY   81   ArgIylleGlnIleHgaSnIleProGlnLeuAsnGlnPglValLeuAspSerLeu 100
DB     |||||
QY   373  TTGGCTCAATATGGGACAGTGGAGAAATGTGGAACAAGTCAACAGACGAAACCGCC 432
DB     |||||
QY   101   LeuAlaGlnIyrGlyThrValGlnAsnCysGlnGlnValAsnThrGlnIleSerGlnThrAla 120
DB     |||||
QY   433  GTTGTCAACGTCACATATGACAAAGAGAGAGCAAAATATGCCATGAGAGAGCTAAGC 492
DB     |||||
QY   121   ValValAsnValThrIyrSerAsnArgGlnGlnThrArgGlnAlaIleMetIylsLeuAsn 140
DB     |||||
QY   493  GGGCATCAGTTTGAAGTACTCCTTCGAAGTTTCTTACATCCCGGATGAAGAGGTGAGC 552
DB     |||||
QY   141   GlnIylsGlnLeuGlnAsnHisAlaLeuIylsSerIyrIleProAspGlnGlnIleThr 160
DB     |||||
QY   553  TCCGCTTCGCCCCCTGACGAGAGCCAGCGTGGGAGACCATCTCCCGGAGCAAGGCCAC 612
DB     |||||
QY   161   -----GlnGlyrProGlnuAsnGlyArgArgIylGlyIylPheGlySerArgIylProArg 178
DB     |||||
QY   613  -----GCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTG 657
DB     |||||
QY   179  GlnGlySerProValAlaAlaGlyAlaProAlaIylsGlnIleProValAspIleProLeu 198
DB     |||||
QY   658  CGGATCCTGGTCCCGCCAGGTTGTGTGGCCATCATCGGAAGAGGGGCTTGACATA 717
DB     |||||
QY   199  ArgLeuLeuValProThrGlnIyrValGlyAlaIleIleGlyLysGlnGlyAlaThrIle 218
DB     |||||
QY   718  AAGAACAATCACTAAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAACTGTGA 777
DB     |||||
QY   219  ArgAsnIleThrIylsGlnThrGlnIleSerIylleAspValHisSarGlyLysGlnuAsnIleGly 238
DB     |||||
QY   778  GCTGCGAGAGAGCTGTACACATTCATGATCCACCCGAGGGGAGCTTGTGAAGCATGGCCG 837
DB     |||||
QY   239  AlaAlaGlnLysAlaIleSerValHisSerThrProGlnGlyCysSerSerAlaCylLys 258
DB     |||||
QY   838  ATGATTTCTGAATTCATGAGGAAGAGGCGAGATGAGCCAACTACGCCGAAGATTCCT 897
DB     |||||
QY   259  MetIleuGlnIylleMetHisIylsGlnAlaIylsAspThrIylsThrAlaAspGlnuAlaPro 278
DB     |||||
QY   898  CTGAAAATCTTGGCACACAAATGCTTGTGTAAGACTGATTGAAAAGAGGACAGAAAT 957
DB     |||||
QY   279  LeuIylleIleuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGlnGlyArgAsn 298
DB     |||||
QY   958  TTGAAGAAATTTGAACATGAAACAGGACCAAGATTAACAATCTCATCTTTGACAGATTG 1017
DB     |||||
QY   299  LeuIylsIylValGlnGlnAspThrGlnThrIylleIleThrIleSerSerLeuGlnAspLeu 318
DB     |||||
QY   1018  AGCATATPACAACCCGGGAAAGAACCATCATCTGTAAGGCGACATTTAGAGGCTGTGCCAGT 1077
DB     |||||
QY   319  ThrLeuIyrAsnProGlnuArgThrIleThrValIylsGlyAlaIleGlnuAsnCysArg 338
DB     |||||
QY   1078  GCTGAGATAGAGATTATGAAAGAGCTGCTGAGGCTTTGAAAATGATATGCTGGCTGTT 1137
DB     |||||
QY   339  AlaGlnGlnGlnIleMetIylsIylValArgGlnuAlaIylGlnuAsnAspValAlaAlaMet 358
DB     |||||
QY   1138  AACACCACTCC-----GGATACCTTC----- 1158
DB     |||||
QY   359  SerLeuGlnIleSerIylleuIleProGlyLeuAsnLeuAlaAlaValGlyLeuProAla 378
DB     |||||
QY   1159  ---TCCAGGCTGTAACCCCATCACAG-----TTTGGCCCGGTCCCGCAT 1200
DB     |||||
QY   379  SerSerSerAlaValAlaProProProSerSerValThrGlyAlaAlaIylProIyrSerSer 398
DB     |||||
QY   1201  CATCACTTTATTCAGAGAGAGAGATTTGAAATCTCTTCAATCCCAACCCAGGCTGGAGC 1260
DB     |||||
QY   399  PheMetGlnAlaIleProGlnGlnIleMetValGlnValPheIleProAlaGlnAlaValGly 418
DB     |||||
QY   1261  GCCATATTCGGGAAAGAGGGGCGACACATCAAAACAGCTGGCGAGATTGCCGAGCTCT 1320
DB     |||||
QY   419  AlaIleIleGlyLysIylsGlyGlnHisIleIylsGlnLeuSerArgPheAlaSerAlaSer 438
DB     |||||
QY   1321  ATCAAGATTGGCCCTGCGGAGAGGCCAGACGTCAAGGAAAGATGTCTATCATCACCGGG 1380
DB     |||||
```


Db 439 IleYstIleAlaProProGluThrProAspSerLysValArgMetValIleThrGly 458
QY 1381 CCACCGGAAGCCGAGTTCAAGGCCCGAGGACGATCTTTGGAAAATGAAAGAGAAAAC 1440
Db 459 ProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysLeuYsGlnGluAsn 478
QY 1441 TTCTTAAACCCCAAGAAAGAGTGAAGTGAAGCGCATACAGAGTCCCTCTCCACA 1500
Db 479 PhePheGlyProLysGlnGlnValLysLeuGlnThrIleArgValProAlaSerAla 498
QY 1501 GCTGGCCGGTGTATGGCAAGAGTGGCAAGACCGTGAACGATCGACAGACTTAACAGT 1560
Db 499 AlaGlyArgValIleGlyLysGlyLysThrValAsnGlnLeuGlnAsnLeuThrAla 518
QY 1561 GCAGAAAGTATCGTGGCTGGTGAACCAAGCCGATGAATAATGAGAAAGTATGTCGA 1620
Db 519 AlaGlnValIleValProArgAspGlnThrProAspGluAsnAspGlnValIleValLys 538
QY 1621 ATTATCGGGCACTCTTTGTAGTCCAGACTGCAAGCCGAGATCAGGGAATTGTACA 1680
Db 539 IleIleGlyAspPheTyrAlaSerGlnMetAlaGlnIleGlyLysIleArgAspIleLeuAla 558
QY 1681 CAGGTGAAGCAGCAGCAGCAGAAATATACCTCAGGAGTGGCTTCACAGCGCAGCAAG 1737
Db 559 GlnValLysGlnGlnIleGlnLysGlyLysIleAsnLeuAlaGlnAlaArgLys 577
RESULT 8
AAB11365
ID AAB11365 standard; protein; 579 AA.
AC AAB11365;
XX 21-FEB-2001 (first entry)
DE Human lung cancer associated antigen L523S.
XX Lung cancer; therapy: treatment; human; tumor; immunogenic; cytostatic;
XX vaccine; detection.
XX Homo sapiens.
XX WO200061612-A2.
XX 19-OCT-2000.
XX 03-APR-2000; 2000WO-US008896.
XX 02-APR-1999; 99US-00285479.
XX 17-DEC-1999; 99US-00466396.
XX 30-DEC-1999; 99US-00476496.
XX 10-JAN-2000; 2000US-00480884.
XX 22-FEB-2000; 2000US-00510376.
XX (CORI-) CORIXA CORP.
XX Wang T, Fan L;
XX MPI; 2000-628399/60.
XX N-PSDB; AAC66035.
PT Isolated polypeptide comprising an immunogenic portion of a lung tumor
PT protein is used for detecting and monitoring progression of lung cancer
PT in a patient.
PS Claim 3; Page 259-261; 261pp; English.
CC This invention describes a novel isolated polypeptide (I) which
CC comprising an immunogenic portion of a lung tumor protein or variant (P2)
CC which have cytostatic activity. The polypeptides and polynucleotides are
CC used in compositions and vaccines to inhibit the development of cancer,
CC especially lung cancer, in a patient. Methods described in the invention
CC can be used to monitor the progression of a cancer by carrying out the
CC detection at subsequent time points and comparing the results from the

CC different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
CC are treated with P2, polynucleotides encoding P2 or antigen presenting
CC cells expressing P2 and then administered to the patient to inhibit
CC development of cancer
XX
XX Sequence 579 AA:
Alignment Scores:
Pred. No.: 5,47e-165 Length: 579
Score: 1875.00 Matches: 369
Percent Similarity: 78.70% Conservative: 78
Best Local Similarity: 64.96% Mismatches: 95
Query Match: 31.86% Indels: 26
DB: Gaps: 6
US-09-270-437D-8 (1-3283) x AAB11365 (1-579)
QY 73 ATGAACAAAGCTTTACATCGGGAACTGAGCCCCCGCCGACCCGAGACCTCCGCGAG 132
Db 1 MetAsnLysLeuTyrIleGlyAsnLeuSerGlnAsnAlaAlaProSerAspLeuGluSer 20
QY 133 CTCTTTGGGACAGGAGAGCTGCCCTGGGAGACAGTCTCTGAAATCGCGGTACGCC 192
Db 21 IlePheLysAspAlaLysIleProValSerGlyProPheLeuValLysThrGlyTyrAla 40
QY 193 TTCTGTGACTACCCCGACCAAGACTGGGCACTCCGCGCATGAGACCCCTCTGGGTAA 252
Db 41 PheValAspCysProAspGluSerThrAlaLeuLysAlaIleGlnAlaLeuSerGlyLys 60
QY 253 GTGCAATGTGCAATGGAAAAATCATGAAATGTGATTCTCAGTCTCTAAAGAGTGAAGC 312
Db 61 IleGlnLeuHISGlyLysProIleGlnValGlnHisSerValProLysArgGlnArgIle 80
QY 313 AGGAATATTCAGATTCGAAACATCCCTCTCAGTCACTGAGTGGAGAGTGTGATGACTT 372
Db 81 ArgLysLeuGlnIleArgAsnIleProPheLysGlnIleProGlnValLeuAspSerLeu 100
QY 373 TTGGCTCAATATGGGACAGTGGAGAAATGTGAAACAGTCAACACAGACAGAAACCGCC 432
Db 101 LeuValGlnTyrGlyValIleGluSerCysGlnGlnValAsnThrAspSerGlnThrAla 120
QY 433 GTTGTCAACGTCACTATGCAACAAGAGAAAGAAATTCAGATGAGAGTGTGATGACTT 492
Db 121 ValIleAsnValThrLysSerSerLysAspGlnAlaArgGlnAlaLeuAspLysLeuAsn 140
QY 493 GGGCATCAGTTTGAAGACTACTCTTCAAGATTCTTCATCCCGATGAAGAGTGAGC 552
Db 141 GlyPheGlnLeuGlnAsnPheThrLeuLysValAlaTyrIleProAspGlnThrAlaIle 160
QY 553 TCCCTTCGCCCCCTCAG-----CGAGCCAGCGT-----GGGACCACTTTCCCGG 600
Db 161 GlnGlnAsnProLeuGlnGlnProArgLysArgArgGlyLeuGlyLysIleArgLysSer 180
QY 601 GAGCAAGGCCACGCCCTTGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGTGGCG 660
Db 181 ArgGlnGly---SerProGlySerValSerLysGlnLysProCysAspLeuProLeuArg 199
QY 661 ATCCGTGCTCCCAACCCAGTTTGTGGCCATCATCGAAAGAGAGGAGCTTGACCATTAAG 720
Db 200 LeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGlnGlyAlaThrIleArg 219
QY 721 AACATCAGTAAGCAGACCGATCCCGGTAGATATTCATGAAAGAAAGAACTGTGAGCT 780
Db 220 AsnIleThrLysGlnThrGlnSerLysIleAspValHisArgLysGlnAsnAlaGlyAla 239
QY 781 GCAGAGAACGCTGTCAACATCCATGCACCCCAAGAGGAGCTTCTGAAGCATCCGATG 840
Db 240 AlaGlnLysSerIleThrIleLeuSerThrProGlnLysThrSerAlaAlaCysLysSer 259
QY 841 ATTCTTGAATCATGAGAAAGAGAGCATGAGACCAACTAGCCGAAGAGATTCTCTG 900
Db 260 IleLeuGlnIleMetHisLysGlnAlaGlnAspIleLysSerThrGlnGlnIleProLeu 279

```
QY 901 AAAATCTGGGACACAAATGGCTTGGTGGAGACTGATTGGAAAAAGAGCAGAAATTTG 960
Db 280 LysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgSerLeu 299
QY 961 AAGAAATTTGAAACATGAAACAGGGGACCAAGATTAACAATCTCATCTTTGGAGATTGGAC 1020
Db 300 LysLysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnGluLeuThr 319
QY 1021 ATATGACACCGGAAAGAAACATCACTGTGAAGGGCACAAGTTGAGCCTGTGGCCAGTCT 1080
Db 320 LeuTyrAsnProGluIleThrIleThrValLysGlyAsnValGluThrCysAlaLysAla 339
QY 1081 GAGATGAGATTATGAAAGAGCTGCGTGAAGGCTTTGAAATGATATGCTGCTGTAAAC 1140
Db 340 GluGluGluIleMetLysLysIleArgGluSerTyrGluAsnAspIleAlaSerMetAsn 359
QY 1141 ACCCACTCC-----GGATACCTTC----- 1158
Db 360 LeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProProThr 379
QY 1159 -----TCCAGCCTGACCCCATCAACCGATTGGCCCTGCCGAT 1200
Db 380 SerGlyMetProProProThrSerGlyProProSerAlaMetThrProProTyrProGln 399
QY 1201 CATCACTCTTATCCAGACAGAGATTGTGATCTTTCATCCCAACCCAGGCTGTGGCC 1260
Db 400 PheGluGln---SerGluThrGluThrValHisLeuPheIleProAlaLeuSerValGly 418
QY 1261 GCCATCATCGGAAAGAAAGGGGCAACATCAACAAGCTGGCGAGATTGGCCGAGCTCT 1320
Db 419 AlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSer 438
QY 1321 ATCAAGATTGGCCCTGGGGAAGGCGGAGCTGACGCAAGAAAGATGTCATCATCAACGGG 1380
Db 439 IleLysIleAlaProAlaGluAlaProAspAlaLysValArgMetValIleIleThrGly 459
QY 1381 CCACCGGAAGCCAGTTCAGAGCCGAGGACGATCTTTGGGAAATGAAAGAGAAAC 1440
Db 459 ProProGluIleGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLysGluGlnAsn 478
QY 1441 TTCTTTAACCCCAAGAAAGATGAAGCTGAGACGCAATCAGAGTCCCTCTTCACA 1500
Db 479 PheValSerProLysGlnGluValLysLeuGlnAlaHisIleArgValProSerPheAla 498
QY 1501 GTGGCGCGGGTGAATGGCAAGGTGGCAGAGCGTGAACGAACTTGAACTTAACAGT 1560
Db 499 AlaGlyArgValIleGlyLysGlyLysThrValAsnGluLeuGlnAsnLeuSerSer 518
QY 1561 GCAGAACTCATCGTGCCTCGTGAACCAACGCGAGATGAAATGAGAGATGATGTCAGA 1620
Db 519 AlaGluValValValProArgAspGlnThrProAspGluAsnAspGlnValValLys 538
QY 1621 ATTATGGGAGCATCTTTGCTTACCGCAGATGCAACGCGCAAGATCAGGAAATTTGACA 1680
Db 539 IleThrGlyHisPheTyrAlaCysGlnValAlaGlnArgLysIleGlnGluIleLeuThr 558
QY 1681 CAGGTGAAGCAGAGAGAGAGAA 1704
Db 559 GlnValLysGlnHisGlnGlnGln 566
```

```
RESULT 9
ABB75053
ID ABB75053 standard; protein; 579 AA.
XX
AC ABB75053;
XX
DT 01-MAY-2002 (first entry)
XX
DE Human lung tumour LS2S3 recombinant protein sequence SPQ ID NO:446.
XX
KW Human; lung tumour; lung cancer; cytostatic; immunostimulant; vaccine;
XX immune response.
XX
```

```
OS Homo sapiens.
XX
EN MO200200174-A2.
XX
XX
XX 03-JAN-2002.
XX
XX 28-JUN-2001; 2001WO-US021065.
XX
XX 28-JUN-2000; 2000US-00606421.
XX
XX 02-AUG-2000; 2000US-00630940.
XX
XX 21-AUG-2000; 2000US-00643597.
XX
XX 15-SEP-2000; 2000US-00662786.
XX
XX 09-OCT-2000; 2000US-00685626.
XX
XX 12-DEC-2000; 2000US-00735705.
XX
XX 07-MAY-2001; 2001US-00850716.
XX
XX (CORI-) CORIXA CORP.
XX
XX Wang T, Wang A, Skeiky YAM, Li SX, Kalos MD, Henderson RA,
XX McNeill PD, Fanger N, Retter MM, Marinakis M, Fanger GR,
XX Vedvick TS, Carter D, Watanabe Y, Peckham DW;
XX
XX WPI; 2002-090513/12.
XX
XX N-PSDB; ABL49297.
XX
XX Polynucleotides encoding lung tumor polypeptides, useful for treating
XX lung cancer or stimulating an immune response.
XX
XX Claim 2; Page 365-367; 374pp; English.
XX
XX The present invention describes human lung tumour proteins. Human lung
XX tumour proteins and polynucleotides have cytostatic and immunostimulant
XX activities, and can be used in vaccine production. Compositions
XX comprising the lung tumour proteins, polynucleotides, antibodies, fusion
XX proteins, T cell populations, or antigen presenting cells that express
XX the lung tumour proteins are useful for treating lung cancer or
XX stimulating an immune response. ABL48959 to ABL49300 and ABB74946 to
XX CC ABL75070 represent sequences used in the exemplification of the present
XX invention
XX
XX Sequence 579 AA;
XX
XX Alignment Scores:
XX
XX Pred. No.: 5,47e-165 Length: 579
XX Score: 1875.00 Matches: 369
XX Percent Similarity: 78.70% Conservative: 78
XX Best Local Similarity: 64.96% Mismatches: 95
XX Query Match: 31.86% Indels: 26
XX DB: 5 Gaps: 6
XX
XX US-09-270-437D-8 (1-3283) x ABB75053 (1-579)
XX
XX 73 ATGAACAAGCTTATCATCGGGAACCTGAGGCGCCGCGTACCGCGCAGACGACTCCGGCAG 132
XX 1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAspLeuGluSer 20
XX
XX 133 CTCTTGGGAGCAGGAAGCTGCCCTGGCGGAGCAGATCCTGCTGAAGTCCGGCTACGCC 192
XX 21 IlePheLysAspAlaLysIleProValSerGlyProPheLeuValLysThrGlyTyrAla 40
XX
XX 193 TTGCGTGACTACCCGCAACGAACTGGGCGATCCGGCCATGAGACACCTCTGGGGTAA 252
XX 41 PheValAspCysProAspSerProIleGluValGluHisSerValProLysArgGlnArgIle 60
XX
XX 253 GTGGAATTCATGAGAAATCATGAAGTTGATTACAGTCTCTAAAAGCTTAAGAGAGC 312
XX 61 IleGluLeuHisGlyLysProIleGluValGluHisSerValProLysArgGlnArgIle 80
XX
XX 313 AAGAAATTCAGATTGAAACATCCCTCTCAACCTGACGAGGAGGATGTTGAGTGAATTT 372
XX 81 ArgLysLeuGlnIleArgAsnIleProProHisLeuGlnIleProLysValLeuAspSerLeu 100
XX
XX 373 TTGGCTCAATATGGGACACTGGAGAAATGTGGAAACAAAGTCAACAGACACAGAAACCGCC 432
```


CC invention
 XX Sequence 579 AA;
 SQ Alignment Scores:
 Pred. No.: 5,47e-165 Length: 579
 Score: 1875.00 Matches: 369
 Percent Similarity: 78.70% Conservative: 78
 Best Local Similarity: 64.96% Mismatches: 95
 Query Match: 31.66% Indels: 26
 DB: 5 Gaps: 6
 US-09-270-437d-8 (1-3283) x ABB74997 (1-579)
 QY 73 ATGAAACAGCTTTTCATCGGGAGACCTGACCCCGCCGTCACCCCGACGACCTCCGCGAG 132
 Db 1 MetasnlyslleuylrlllelyasnleuserglunsnalalalaProserAspSerleu 20
 QY 133 CTCTTTGGGGACAGAAAGCTGCCCCGTGGCGGACAGAGCTCTGTGAAGTCCGGCTACGCC 192
 Db 21 llepelysaspalalysileProvalSerGlyProheleuvallysrthglyrAla 40
 QY 193 TTGCTGACATAACCCGACACAGAACTGGGCCATCCGCGCCATCGAACCCCTCTCGAGTAA 252
 Db 41 PheValaspCysProaspGlnSerTrpAlaleuylsAlalleglunalaleuSerGlyls 60
 QY 253 GTGGAAATTCAGTGGGAAATCATCGAAGTTGATTACTCAGCTCTAAAGCTTAAGAGC 312
 Db 61 lleglunelhseglylrsProillegluvalglunhservalProlysArglnArgile 80
 QY 313 AGGAAATTCAGATTGGAACATCCCTCTCACCCTGAGTGGAGAGTGTGGATGACTT 372
 Db 81 ArglylsleuglnlleargasnilleProprohlsleuclntrpgluvalleuaspSerleu 100
 QY 373 TTGGCTCAATATGGGACAGTGGAGATGTGAAACAGTCAACACAGACAGAAACCGCC 432
 Db 101 leuvalglntlylglvalValgluserCysgluglnValasnThrAspserglunThrAla 120
 QY 433 GTTGTCAACGTCACATTCGACACAGAGAGAGCAAAATAGCCATGAGAGACTAGC 492
 Db 121 ValValasnvalThrlyrSerSerlySaspGlnAlaArglnAlaleuaspLySleasn 140
 QY 493 GGGCATCATGTTTGAAGACTACTCCTTCAGATTTCCTACATCCCGGATGAAGAGTGGAGC 552
 Db 141 GlyPheglunleuglunsnphethrleuylValAlatylrilleProaspGlnThrAlaAla 160
 QY 553 TCCCTTCGCCCCCTAG-----CGAGCCACGCT-----GGGACCACTTCCCGG 600
 Db 161 GlnGlnasnProleuglnInpProArgGlyArgArgGlyleugllyGlnArgGlySerSer 180
 QY 601 GAGCAAGGCCACGCCCTGGGGGACACTTCTCAGGCCACAGACAGATTGATTTCCCGCTGGG 660
 Db 181 ArgGlnGly---SerProGlySerValSerlySglnlySerProCysAspLeuProleuArg 199
 QY 661 ATCTGGTCCCCCAACCATGTTTGTGGTCATCATCGGAAAGAGGGCTTGACCATTAAG 720
 Db 200 leuLeuValProthrGlnPheValGlyAlallellelylysglnGlyAlaThrIleArg 219
 QY 721 AACATACATAAGACACCAAGTCCGGAGTAGATATCCATGAAGAAAGAACTCTGAGCT 780
 Db 220 AsnIleThrlysglnThrGlnSerlySileAspValHlsArglysglnunsnalaglyAla 239
 QY 781 GCAGAGAACCTGTGACCATCATCGACCCCAAGAGGGAGCTTCTGAAGATCCCGCAG 840
 Db 240 AlaGlnlySerIleThrIleleuSerThrProGlnGlyThrSerAlaIaCyslySer 259
 QY 841 ATTCTTGAATCATGAGAAAGAGGACAGATGAGACCAAACTAGCCAGAGATTCTCTCG 900
 Db 260 lleleuglnlleMerHislysglnAlaGlnAspIlelyPheThrGlnGlnleProleu 279
 QY 901 AAAATTTGGCACACATGCTGTTGTTGAAGACTGATTGGAAGAAAGACAGAAATTTG 960
 Db 280 LysIleleuAlaHisasnAsnPhValGlyArgleuIleGlylysglnGlyArgsnleu 299

QY 961 AAGAAATTTGAACATGAAAAAGGACCAAGATAACATCTCATCTTTCAGAGATTGGAGC 1020
 Db 300 LyslyslleGlnInAspThrAspThrlysrIleThrIleSerProleuGlnGlnleuThr 319
 QY 1021 ATATACAAACCCGGAAGAAACCATCATCTGTGAAGGACACAGTGTGAGGCTGTGCCAGTCT 1080
 Db 320 leuTrpAsnProGlnArgThrIleThrVallysglyAsnValGlnThrCysAlaIysAla 339
 QY 1081 GAGATAGAGATTATGAAAGACCTGCGAGGCTTGTGAATAATGATATGCTGCTGTAC 1140
 Db 340 GlnGlnGlnIleMetlylysrIleArgGlnSerlyrGlnAsnAspIleAlaSerMetasn 359
 QY 1141 ACCCACTCC-----GGAATACCTC----- 1158
 Db 360 leuGlnAlaHlsleuIleProGlyleuAsnleuSnAlaleuGlyleuPheProbrothr 379
 QY 1159 -----TCAGCTGTACCCCATCATCACAGTTTGGCCGCTTCCCGCAT 1200
 Db 380 SerGlyMetProProProthrSerGlyProProserAlaMetThrProProTyProGln 399
 QY 1201 CATCACTCTTATCCAGACAGAGAGATGTGATCTTCAATCCCAACCCAGCGTGGGC 1260
 Db 400 PheGlnGln---SerGlnThrGlnThrValHlsleuPheIleProAlaleuSerValGly 418
 QY 1261 GCCATCATCGGGAAGAAAGGCGGACACATCAAAACAGCTGCGAGATTCGCGAGCCTT 1320
 Db 419 AlaIlelleGlylysglnGlnHlsIlelyGlnleuSerArgPheAlaGlyAlaSer 438
 QY 1321 ATCAGATTGCCCCGTGGGGAAGGCCACAGACTCAGCAAGAAAGATGTCATATATACCGGG 1380
 Db 439 IlelyslleAlaProAlaGlnAlaProAspAlaIysValArgMetValIleIleThrGly 458
 QY 1381 CCACCGGAAGCCAGTTTCAAGGCCACAGGACGAGTCTTGTGGAAACGAAAGAGAAAC 1440
 Db 459 ProProGlnAlaGlnPheylsAlaGlnGlyArgIlelyrGlylysrIlelySglnGlnasn 478
 QY 1441 TTCTTTAACCCCAAGAAAGTGAAGCTTGAAGCGCATATCAGAGTGCCTTCCACA 1500
 Db 479 PheValSerProlysglnGlnVallySleuglnAlaHlsIleArgValProserPheAla 498
 QY 1501 GCTGGCGCGGTTGTTTGAAGAGTGGCAGACCGTGAACCAATGCGAACTTAACAGT 1560
 Db 499 AlaGlyArgValIleGlylysglyGlylysrThrValAsnGlnleuGlnAsnleuSerSer 518
 QY 1561 GCAGAGTATCGTGCCTGATGACCAACGCCAGATGAAGAAATGAGAGATGCTGACA 1620
 Db 519 AlaGlnValValAlaProArgAspGlnThrProAspGlnAsnAspGlnValValIys 538
 QY 1621 ATTATCGGCACTTCTTCTAGCCAGACTGACACAGCGCAAGATCAGAAATGTACAA 1680
 Db 539 IleThrGlyHlsPheThrAlaCysGlnValAlaGlnAlaArglyslleGlnGlnIleleuThr 558
 QY 1681 CAGGTGAACAGACAGAGACGAGAA 1704
 Db 559 GlnVallysglnHlsGlnGlnGln 566
 RESULT 11
 ABB75054
 ID ABB75054 standard, protein; 579 AA.
 XX ABB75054;
 AC
 XX
 DT 01-MAY-2002 (first entry)
 XX
 DE Human lung tumour L523S recombinant protein sequence SEQ ID NO:449.
 KW Human; lung tumour; lung cancer; cytostatic; immunostimulant; vaccine;
 KW immune response.
 XX
 OS Homo sapiens.
 XX
 PN MO200200174-A2.

Db 479 PheValSerProIysGlnValIysLeuGlnAlaHisIleArgValProSerPheAla 498
QY 1501 GCTGGCCGGGATTGGCAAAAGTGGCAGACCGTGAACCTGCAAGACTTAACCACT 1560
Db 499 AlaGlyArgValIleGlyIysGlyIysThrValAsnGlnLeuGlnAsnLeuSer 518
QY 1561 GCAGAAATGATGTCGCTGCTGACCAAAAGCCAGATGAAATAGGAATGATGCTCAGA 1620
Db 519 AlaGlnValValProArgAspGlnThrProAspGlnAsnAspGlnValValIys 538
QY 1621 ATTATCCGGCACTTCTTGTCTAGCCAGACTGCACAGCCGCAAGATAGAGAAATGTACAA 1680
Db 539 IleThrGlyHisPheTyrAlaCysGlnValAlaGlnArgIysIleGlnGlnIleuLeuThr 558
QY 1681 CAGTGAAGCAGCAGCAGCAAAA 1704
Db 559 GlnValIysGlnHisGlnGlnGln 566
RESULT 12
ABP61917
ID ABP61917 standard; protein; 579 AA.
XX
AC ABP61917;
XX
DT 07-OCT-2002 (first entry)
XX
DE Human lung cancer associated protein sequence SEQ ID NO:348.
XX
KW Human; lung cancer; lung tumour; cytostatic; gene therapy; vaccine.
XX
OS Homo sapiens.
XX
PN W0200247534-A2.
XX
PD 20-JUN-2002.
XX
PF 30-NOV-2001; 2001WO-US047576.
XX
PR 12-DEC-2000; 2000US-00735705.
PR 07-MAY-2001; 2001US-00850716.
PR 28-JUN-2001; 2001US-00897778.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Wang A, Skeiky YAW, Li SX, Kalos MD, Henderson RA,
PI Menelli PD, Fanger N, Retter MW, Durham M, Fanger GR, Vedvick TS;
PI Carter D, Watanabe Y, Peckham DW, Cai F, Foy TM;
XX
XX WPI; 2002-583465/62.
XX
XX N-PSDB; ABQ92440.
XX
PT Novel lung carcinoma polynucleotide sequences and polypeptides encoded by
PT the polynucleotides, useful in pharmaceutical compositions such as
PT vaccines and as markers to indicate the presence of lung cancer.
XX
XX Example 2; Page 337-339; 381pp; English.
XX
CC The present invention describes isolated human lung carcinoma
CC polynucleotides (I) and polypeptides (II). (I) and (II) have cytostatic
CC activity, and can be used in gene therapy and in vaccines. Compositions
CC comprising (I) or (II) can be used for stimulating an immune response in
CC a patient and for treating lung cancer in a patient. Oligonucleotides of
CC (I) can be used for detecting lung cancer in a patient. Oligonucleotides of
CC obtaining a biological sample from the patient, contacting the biological
CC sample with the oligonucleotide, detecting in the sample, an amount of
CC polynucleotide that hybridizes to the oligonucleotide and comparing the
CC amount of polynucleotide that hybridizes to the oligonucleotide to a
CC predetermined cut-off value, and determining the presence of a cancer in
CC the patient. (I) and (II) are useful in pharmaceutical compositions, e.g.
CC vaccines. (I) is useful as a marker to indicate the presence or absence
CC of a cancer such as lung cancer. ABQ92145 to ABQ92486 and ABP61866 to
CC ABP61992 represent sequences used in the exemplification of the present

CC invention
XX
SQ Sequence 579 AA;
Alignment Scores:
Pred. No.: 5,47e-165 Length: 579
Score: 1875.00 Matches: 369
Percent Similarity: 78.70% Conservative: 78
Best Local Similarity: 64.96% Mismatches: 95
Query Match: 31.86% Indels: 26
DB: 5 Gaps: 6
US-09-270-437D-8 (1-3283) x ABP61917 (1-579)
QY 73 ATGAACAAGCTTTACATCGGGAACCTGAGCCCGCCGTCACCGCAGACCTCCGGGCA 132
Db 1 MetAsnIysLeuTyrIleGlyAsnLeuSerGlnAsnAlaAlaProSerAspLeuIleuSer 20
QY 133 CTCTTTGGGGACAGGAAGCTGCCCTGGCGGGACAGGTCTGCTGAAGTCCGGCTAGCGC 192
Db 21 IlePheIysAspAlaIysIleProValSerGlyProPheLeuValIysThrGlyTyrAla 40
QY 193 TTGCTGGACTACCCCGACAGACACTGGCGGCATCCGCGCATCGAGACCCCTCCGGGTAA 252
Db 41 PheValAspCysProAspGlnSerTrpAlaLeuIysAlaIleGlnAlaLeuSerGlyIys 60
QY 253 GTGAATTCATGCGAAATATCATGAAATTGATTACTGCTCTTAAAGCTAAGAGAGC 312
Db 61 IleGlnLeuHisGlyIysProIleGlnValGlnHisSerValProIysArgGlnArgIle 80
QY 313 AGGAAATTCAGATTTCGAAACATCCCTCTCTACCTGCATGCGGAGGTGTGATGACTT 372
Db 81 ArgIysLeuGlnIleArgAsnIleProHisIleuGlnTrpGlnValLeuAspSerLeu 100
QY 373 TTGGCTCAATATGAGGACAGTGGAGATGGGAAACATGCAACAGACACAGCAACCGCC 432
Db 101 LeuValGlnTyrGlyValValGlnSerCysGlnLeuIleAsnTrpAspSerGlnThrAla 120
QY 433 GTTGTCACGCTCATATATGCAACAGAAAGCAAAATAGCCATGAGAGACTAAGC 492
Db 121 ValValAsnValThrTyrSerSerIysAspGlnAlaArgGlnAlaLeuAspIysLeuAsn 140
QY 493 GGGCATGAGTTTGAAGACTACTCCTTCAAGATTTCCATATCCCGATGAAAGGTGAGC 552
Db 141 GlyPheGlnLeuGlnAsnPheThrLeuIysValAlaTyrIleProAspGlnThrAlaAla 160
QY 553 TCCCTTGGCCCGCTAG-----CGAGCCCAAGCT-----GGGACCACTTCCCGG 600
Db 161 GlnGlnAsnProLeuGlnGlnProArgGlyArgArgGlyLeuGlnArgIysSer 180
QY 601 GACCAAGCCACGCGCCCTGGGGCACTTCTCAGCGCCAGACAGATTGATTCGCGTGGG 660
Db 181 ArgGlnGly--SerProGlySerValSerIysGlnIysProCysAspLeuProLeuArg 199
QY 661 ATCTGTGTCCTCCCAAGTTGTTGGTGCATATATGGAAGAGAGGCTTGACCATTAAG 720
Db 200 LeuLeuValProThrGlnPheValGlyAlaIleIleGlyIysGlnGlyAlaThrIleArg 219
QY 721 AACATCACTAAGAGAGACCCAGTCCGGGTAGATATCATATGAAAGAAAGAACTCTGAGCT 780
Db 220 AsnIleThrIysGlnThrGlnSerIysIleAspValHisArgIysGlnAsnAlaGlyAla 239
QY 781 GCAGAGAACCTGTCACTCATTCATGACCCAGAGAGGAGACTTGAAGATCCGGCATG 840
Db 240 AlaGlnIysSerIleThrIleLeuSerThrProGlnGlyThrSerAlaIaCysIysSer 259
QY 841 ATTCTTGAATATCATGAGAAAGAGCGATGAGACCAATCAGCCGAGAGATTCTCTG 900
Db 260 IleLeuGlnIleMetHisIysGlnIaGlnAspIleIysPheThrGlnIleProLeu 279
QY 901 AAAATCTTGGCACACAAATGCTTGGTGGAAACATGATTGAAAGAAAGAGCAATTTG 960
Db 280 IysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyIysGlnGlyArgAsnLeu 299

```

QY 961 AAGAAATTTGAACATGAACAGGAGCCAGATTAACATTCATCTTTGACAGATTGAGC 1020
DB 300 LyleylsilegluglnasptlrAspThrLysleThrleSerProleuglnleuThr 319
QY 1021 ATATACAAACCCGAAACCATCATCTGTGAAGGCAAGTTGAGGCGCTGTGACGTCT 1080
DB 320 LeutyrasnProgluAthrThrleThrValysleValnGlnThrCysAlaLysAla 339
QY 1081 AGATATAGATTTATGAAGAGCTGCTGAGCCCTTTGAAAAATGATGATGCTGTTAAC 1140
DB 340 GluGluGlnleuMetLysleValArgLysleuThrValnAspIleAlaSerMetasn 359
QY 1141 ACCCACTCC-----GGATCTTC----- 1158
DB 360 LeuGlnAlaHisleuLeuProGlyLeuasnleuasnAlaLeuGlyLeuPheProProthr 379
QY 1159 -----TCCAGCCTGTACCCCATCACAGTTTGCGCCCTCCCGCAT 1200
DB 380 SerGlyMetProProProThrSerGlyProProSerAlaMetThrProProTyProGln 399
QY 1201 CATCACTCTTATCCAGACAGAGATTGTGATCTTTTCATCCCAACCCAGGCTGTGGCC 1260
DB 400 PheGluGln-----SerGluThrGlnThrValHisleuPheIleProAlaLeuSerValGly 418
QY 1261 GCCATCATCGGAGAGAGGCGGCACTCAATCAACAGCTGGCGAGATTGGCGGAGCCCT 1320
DB 419 AlaIleIleGlyLysGlnIleGlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSer 438
QY 1321 ATCAAGATTGCCCCCGGAGAGCCCAAGCGTCAGCGAAAGGATGTCATCATCACCGG 1380
DB 439 IleLysIleAlaProAlaGlnAlaProaspAlaLysValArgMetValIleIleThrGly 458
QY 1381 CCACCGGAGCCCGAGTTGAGGCGCCAGAGCGGATCTTTGGAAAAGTGAAGAGGAAAGC 1440
DB 459 ProProGlnAlaGlnPheLysAlaGlnIleArgIleTyGlyLysIleLysGlnGlnasn 478
QY 1441 TTCTTTAACCCTAAAGAAAGATGAAGCTGGAAGCGCAATATGAGAGTGCCTTCACCA 1500
DB 479 PheValSerProLysGlnIleValLysleuGlnAlaHisIleArgAlaProSerPheAla 498
QY 1501 GGTGCGCGGGTATTGGCAAGGTGCGCAAGCGTGAACGAATCGAAGATTAAACAGT 1560
DB 499 AlaGlyAlaValIleGlyLysGlyLysThrValAsnGlnleuGlnasnleuSerSer 518
QY 1561 GCAGAAGTCATCGTCTGCTGTGACCAACGCCAGATGAATAATGAGGAAGTATCTCA 1620
DB 519 AlaGlnValValValProArgAspGlnThrProAspGlnAsnAspGlnValValLys 538
QY 1621 ATTATCGGCACTTTCTTTGTAGCCCAAGTCTGCAACCGCAAGTTCAGGAAATGTACA 1680
DB 539 IleThrGlyHisPheTyraLysGlnValAlaGlnArgLysIleGlnGlnIleleuThr 558
QY 1681 CAGGTGAAGCAGCAGGAGCAAAA 1704
DB 559 GlnValLysGlnHisGlnGlnGln 566

```

RESULT 13
ABP61974 standard; protein; 579 AA.

```

XX AC ABP61974;
XX DT 07-OCT-2002 (first entry)
XX DE Human lung cancer associated protein sequence SEQ ID NO:449.
XX KW Human; lung cancer; lung tumour; cyrostatic; gene therapy; vaccine.
XX OS Homo sapiens.
XX PN MO200247534-A2.

```

```

PD 20-JUN-2002.
PE 30-NOV-2001; 2001WO-US047576.
XX 12-DEC-2000; 2000US-00735705.
PR 07-MAY-2001; 2001US-00850716.
PR 28-JUN-2001; 2001US-00897778.
XX (CORI-) CORIXA CORP.
XX Wang T, Wang A, Skeiky YAW, Li SX, Kalos MD, Henderson RA;
PI McNeil PD, Fanger N, Retter MW, Durham M, Fanger GR, Vedvick TS;
PI Carter D, Watanabe Y, Peckham DW, Cai F, Foy TM;
XX WPI; 2002-583465/62.
DR N-PSDB; ABQ92485.
XX
PT Novel lung carcinoma polynucleotide sequences and polypeptides encoded by
PT the polynucleotides, useful in pharmaceutical compositions such as
PT vaccines and as markers to indicate the presence of lung cancer.
PS Claim 9; Page 375-377; 381pp; English.
XX
CC The present invention describes isolated human lung carcinoma
CC polynucleotides (I) and polypeptides (II). (I) and (II) have cyrostatic
CC activity, and can be used in gene therapy and in vaccines. Compositions
CC comprising (I) or (II) can be used for stimulating an immune response in
CC a patient and for treating lung cancer in a patient. Oligonucleotides of
CC (I) can be used for detecting the presence of a cancer in a patient, by
CC obtaining a biological sample from the patient, contacting the biological
CC sample with the oligonucleotide, detecting in the sample, an amount of
CC polynucleotide that hybridises to the oligonucleotide and comparing the
CC amount of polynucleotide that hybridises to the oligonucleotide to a
CC predetermined cut-off value, and determining the presence of a cancer in
CC the patient. (I) and (II) are useful in pharmaceutical compositions, e.g.
CC vaccines. (I) is useful as a marker to indicate the presence or absence
CC of a cancer such as lung cancer. ABQ92145 to ABQ92486 and ABP61866 to
CC ABP61992 represent sequences used in the exemplification of the present
CC invention
XX
SQ Sequence 579 AA:

```

Alignment Scores:
Pred. No.: 5,47e-165 Length: 579
Score: 1875.00 Matches: 369
Percent Similarity: 78.70% Conservative: 78
Best Local Similarity: 64.96% Mismatches: 95
Query Match: 31.86% Indels: 26
DB: Gaps: 6

US-09-270-437d-8 (1-3283) x ABP61974 (1-579)

```

QY 73 ATGAACACAGCTTATACATCGGAGAACCTGAGCCCGCTGACCGCAGACCTCCGGCAG 132
DB 1 MetAsnLysleuTyrlleGlyAsnleuSerGlnAsnAlaIleProSerAspLeuGlnSer 20
QY 133 CTCTTTGGGAGCAGGAAGTCCCTGCGGCGGACAGGCTCTGCTGAAGTCCGCTACGCC 192
DB 21 IlePheLysAspAlaLysIleProValSerGlyProPheLeuValLysThrGlyTyraL 40
QY 193 TTGCGTACTACCCCGACCGAAGCTGCGGCGGCACTGCGGCGGCACTGCGGCTGAAA 252
DB 41 PheValAspCysProaspGlnSerIlePheValAlaIleGlnAlaLeuSerGlyLys 60
QY 253 GTGGAATTGCAATGGGAAATCATGAGATTGATTAAGTCTTAAAGTAAAGTAAAGAGC 312
DB 61 IleGlnleuHisGlyLysProIleGlnValGlnHisSerValProLysArgGlnArgIle 80
QY 313 AGGAAATTCAGATTGCAACATCCCTCTCACTGAGGAGGAGGAGGAGGAGGAGGAGCTT 372
DB 81 ArgLysleuGlnIleArgAsnIleProProHisleuGlnIleArgGlnValLeuAspSerleu 100
QY 373 TTGGCTCAATATGAGGACAGTGGAGAAATGTGGAACAAGTCAACACAGAACAGGCC 432

```


Db	101	LeuValGlnIrrGelValValIgluSerCysGlnGlnValAsnThrSpsSerGlnThrAla	120
QY	433	GTGTCAAGCTCACTATGCAACAGAGAAAGAAAATTCATTCAGAGAGCTTAAC	492
Db	121	ValValAsnValThrTyrSerSerLysAspGlnAlaArgGlnAlaLeuSpsLysAsn	140
QY	493	GGGCATCAGTTTAGAACTACTCTTCAGATTTCCTACATCCCGGATGAAGAGTGAAC	552
Db	141	GlyPheGlnLeuGlnAsnPheThrLeuLysValAlaIleTyrIleProAspGlnThrAlaIa	160
QY	553	TCCCTTCGGCCCCCTGAG-----GGAGGCCAGGCT-----GGGACCACTTCCTCCGG	600
Db	161	GlnGlnAsnProLeuGlnGlnIleProAlaArgGlyAlaArgGlyLeuGlyGlnArgLysSer	180
QY	601	GAGCAAGGCCACGCCCTCCGGGACATTCCTCAAGGCCACAGATTCATTCCTCCGCTGGG	660
Db	181	ArgGlnGly---SerProGlySerValSerLysGlnLysProCysAspSpsLeuProLeuArg	199
QY	661	ATCTCGTCCCAACCCAGTTGTGTGGTCATCATCGAAAGAGGGCTTGACCAATAAG	720
Db	200	LeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGlnGlyAlaThrIleArg	219
QY	721	AACATACATAAGAGAGCCAGTCCCGGCTAGATATTCATATGAAAAGAACTCTGAGCT	780
Db	220	AsnIleThrLysGlnThrGlnSerLysIleAspValIleArgLysGlnAsnIaGlyAla	239
QY	781	GCAGAGAAAGCTGTGCACATTCATGCACACCCAGAGGGGACATTCGAAAGCATGCCGAC	840
Db	240	AlaGlnLysSerIleThrIleLeuSerThrProGlnGlyThrSerAlaIaLysLysSer	259
QY	841	ATTCTTGAATCATGTCAGAAAGAGCGAGATGAGACCACTGACCGAAGAGATCTCTGTG	900
Db	260	IleLeuGlnIleMetCHisLysGlnIaGlnAspIleLysPheThrGlnGlnIleProLeu	279
QY	901	AAATCTTGGGACACAAATGCGTGTGGTTGGAAGCTGATTCGAAAAGAAAGCAAAATTGG	960
Db	280	LysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGlnGlyArgAsnLeu	299
QY	961	AAGAAATTTGAACATGAAACAGAGGACCAAGATTAACATCTGATCTTGCAGAGATTTGAGC	1020
Db	300	LysLysIleGlnGlnAspThrAspThrLysIleThrIleSerProLeuGlnGlnLeuThr	319
QY	1021	ATATTCAAACCCGAAAGAACCATCATCTGTGAAGGGCACAGTTGAGGCTGTGCAGTGTCT	1080
Db	320	LeuTyrAsnProGlnArgThrIleThrValLysGlnAsnValGlnThrCysAlaLysAla	339
QY	1081	GAGATAGAGATTATGAAAGAGCTGGGTGAGGCTTTGAAAATGATATGCTGAGCTGTAAAC	1140
Db	340	GlnGlnGlnIleMetLysLysIleArgGlnSerTyrGlnAsnAspIleAlaSerMetAsn	359
QY	1141	ACCCATGCC-----GGATACCTTG-----GGATACCTTG-----GGATACCTTG-----	1158
Db	360	LeuGlnAlaHisLeuIleIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProThr	379
QY	1159	-----TCCAGCCTGTACCCCATCAACAGTTTGCCCGCTTCCCGCAT	1200
Db	380	SerGlyMetProProThrSerGlyProProSerAlaMetThrProProTyrProGln	399
QY	1201	CATCACTCTTATCCAGAGCAGAGAAATTGTGAATCTTTCATCCCAACCCAGGCTGTGGC	1260
Db	400	PheGlnGln---SerGlnThrGlnThrValHisLeuPheIleProAlaLeuSerValGly	418
QY	1261	GCCATCATCGGAGAGAGAGGGGACACATCAAAACAGCTGGGAGATTCGCGCGAGCTCT	1320
Db	419	AlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSer	438
QY	1321	ATCAAGATTGCCCTGCGGAGAGGCCACAGCTGACGAAAGATGTCATCATCAACGGG	1380
Db	439	IleLysIleAlaProIaGlnAlaProAspAlaLysValArgMetValIleIleThrGly	458
QY	1381	CCACCGGAGCGGCTTCAAGGCCACAGGACGATCTTTGGGAAACTGAAAGAGGAAAC	1440

Db	459	ProProGluAlaGlnPheValAlaGlnGlyArgIleTyrGlyLysIleLysGluGluAsn	478
QY	1441	TTCTTTAAACCCCAAGAAAGAAAGTGAAGCTGGAAAGCGATATACAGAGTGGCCCTCTTCACA	1500
Db	479	PheValSerProLysGluGluValIleLysIleuGluAlaHisIleArgValProSerPheAla	498
QY	1501	GCTGGCGGGTGATTTGGCAAGGTCGCAAGCCGTGAACCACTGCAGACTTAAACAGT	1560
Db	499	AlaGlyArgValIleGlyLysGlyLysThrValAsnGluLeuGlnAsnLeuSerSer	518
QY	1561	GCAGAAAGTCATCGTGGCTCGTGAACCAACGCCATGATGAAAATAGAGGAAGTGATGTCACA	1620
Db	519	AlaGluValValValProArgAspGlnThrProAspGluAsnAspGlnValValLys	538
QY	1621	ATTATCGGCGACTCTTTTGTCTAGCCAGACTGCACAGCGCAAGATCAGGAAATTGTACAA	1680
Db	539	IleThrGlyHisPheTyrAlaCysGlnValAlaGlnArgLysIleGlnGluIleLeuThr	558
QY	1681	CAGGTGAAGCAGCAGCAGCAGCAAA 1704	
Db	559	GlnValLysGlnHisGlnGlnGln 566	
RESULT 14			
ABP61973	ID	ABP61973 standard; protein; 579 AA.	
XX	AC	ABP61973;	
XX	DT	07-OCT-2002 (first entry)	
XX	DE	Human lung cancer associated protein sequence SEQ ID NO:446.	
XX	KM	Human; lung cancer; lung tumour; cytostatic; gene therapy; vaccine.	
OS	XX	Homo sapiens.	
XX	XX	WO200247534-A2.	
PN	XX	20-JUN-2002.	
PD	XX	30-NOV-2001; 2001MO-US047576.	
XX	XX	12-DEC-2000; 2000US-00735705.	
PR	XX	07-MAY-2001; 2001US-00850716.	
PR	XX	28-JUN-2001; 2001US-00897778.	
PA	XX	(CORI-) CORIXA CORP.	
XX	XX	Wang T, Wang A, Skeiky YAM, Li SX, Kalos MD, Henderson RA;	
PI	PI	McNeill PD, Fanger N, Retter MW, Durham M, Fanger GR, Vedvick TS;	
PI	PI	Carter D, Watanabe Y, Peckham DW, Cai F, Foy TW;	
DR	DR	WPI; 2002-583465/62.	
DR	DR	N-PSDB; ABQ92483.	
XX	XX	Claim 9; Page 372-374; 381pp; English.	
CC	CC	The present invention describes isolated human lung carcinoma	
CC	CC	polynucleotides (I) and polypeptides (II). (I) and (II) have cytostatic	
CC	CC	activity, and can be used in gene therapy and in vaccines. Compositions	
CC	CC	comprising (I) or (II) can be used for stimulating an immune response in	
CC	CC	a patient and for treating lung cancer in a patient. Oligonucleotides of	
CC	CC	(I) can be used for detecting the presence of a cancer in a patient, by	
CC	CC	obtaining a biological sample from the patient, contacting the biological	
CC	CC	sample with the oligonucleotide, detecting in the sample, an amount of	
CC	CC	polynucleotide that hybridises to the oligonucleotide and comparing the	
CC	CC	amount of polynucleotide that hybridises to the oligonucleotide to a	
CC	CC	predetermined cut-off value, and determining the presence of a cancer in	
CC	CC	the patient. (I) and (II) are useful in pharmaceutical compositions.	

CC vaccines. (1) is useful as a marker to indicate the presence or absence
 CC of a cancer such as lung cancer. AB092145 to AB092486 and ABP61866 to
 CC ABP61992 represent sequences used in the exemplification of the present
 CC invention

XX Sequence 579 AA:

Alignment Scores:

Pred. No.:	5,47e-165	Length:	579
Score:	1875.00	Matches:	369
Percent Similarity:	78.70%	Conservative:	78
Best Local Similarity:	64.96%	Mismatches:	95
Query Match:	31.86%	Indels:	26
DB:	5	Gaps:	6

US-09-270-437d-8 (1-3283) x ABP61973 (1-579)

```

QY 73 ATGAACAGCTTTATCATCGGGAACTGAGCCCGCTCACCGCCGAGACCTCCGGCAG 132
DB 1 MetasnllyleuTrilleglyasnleusergluamalaalProserAspleuGluser 20
QY 133 CTCTTTGGGACAGAGAGCTGCCCTGGCGGAGACGTCCTGTAAGTCCGGCTACGCC 192
DB 21 IlePheLysAspAlaLysleleProValSerGlyProPheLeuValLysThrGlyTrala 40
QY 193 TTCGTGAGTACCCCGACCGAGAACTGGGCGCATCCGGGCATGAGACCCCTCTCGGATAA 252
DB 41 PheValAspCysProAspGluSerTrpAlaLeuLysAlaAlaLeuLysThrGlyLys 60
QY 253 GTGGAATTCATGGGAAATCATCGAAGTGTATCTCATGCTCTAATAAAAGTAAAGAGC 312
DB 61 IleGluLeuHisGlyLysProIleGluValGluHisSerValProLysArgGlnArgIle 80
QY 313 AGCAAAATTCAGATTGGAACATCCCTCTCATCTGCACTGGAGGTTGTGATGACTT 372
DB 81 ArgLysLeuGlnIleArgAsnIleProPheHisLeuGlnTrpGluValLeuAspSerleu 100
QY 373 TTGGCTCAATTATGGGACATGAGATGTGGAACATGCAACACACAGAAACCGCC 432
DB 101 LeuValGlnTrpGlyValValGluSerCysGluGlnValAsnThrAspSerGluThrAla 120
QY 433 GTTGTACAGCTCATATGCAACACAGAAAGCAAAATAGCCATGAGAAAGCTAAC 492
DB 121 ValValAsnValThrLysSerSerLysAspGlnAlaArgGlnAlaLeuAspLysleuAsn 140
QY 493 GGGCATCAGTTGAGAACTACTCTTCATGATTTCTCACTCCCGATGAAAGGTGAGC 552
DB 141 GlyPheGlnLeuGluAsnPheThrLeuLysValAlaLysTrpProAspGluThrAlaAla 160
QY 553 TCCCTTCGCGCCCTCAG-----CGAGCCGAGGT-----GGGAGCACTCTTCCCG 600
DB 161 GlnGluAsnProLeuGlnGlnProArgGlyArgArgGlyLeuGluGlnArgLysSer 180
QY 601 GAGCAAGGCGACCGCTCGGGGCACTTTCAGGCGCAGACAGATGATTTCCGCTGCGG 660
DB 181 ArgGlnGly---SerProGlySerValSerLysGlnLysProCysAspLeuProLeuArg 199
QY 661 ATCTGTGTCACCCACCGATTGTTGTGTCATCATCGGAAAGAGGCTTGACATTAAG 720
DB 200 LeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGlnGlyAlaThrIleArg 219
QY 721 AACATCCTAAGAGACACCGCTCCGGGTAGATTCATGATGAAAGAAAGAACTCGAGCT 780
DB 220 AsnIleThrLysGlnThrGlnSerLysIleAspValHisArgLysGluAsnAlaGlyAla 239
QY 781 GCAGAGAGCGCTGCACCATCATGACACCCAGAGGAGACTTGTGAAGCATGCGCATG 840
DB 240 AlaGluLysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCysLysSer 259
QY 841 ATTCTTGAATCATGCAAGAAAGAGCGAGATGAGACMAAATCAGCCGAAAGATTCCTCTG 900
DB 260 IleLeuGluIleMetHisLysGluAlaGlnAspIleLysPheThrGlnGluIleProLeu 279

```

```

QY 901 AAAATCTTGACACAATAGCGTTGGTGGAAAGACTGATGGAAAAAGAGCGAAATTTG 960
DB 280 LysIleLeuValHisSerAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeu 299
QY 961 AAGAAATTTGAACTGAAACAGGAGCAAGATTAATCTCATCTTTGACAGATTTGAGC 1020
DB 300 LysLysIleGluGlnAspThrAspThrLysIleThrLysSerProLeuGlnLeuThr 319
QY 1021 ATATACAAACCGGAAAGAACCATCATCTGTGAAGGCGACAGTGTAGGCTGTGCCAGTCT 1080
DB 320 LeuTrpAsnProGluValArgThrIleThrValLysGluAsnValGluThrCysAlaLysAla 339
QY 1081 GAGATGAGATTAATGAGACAGCTGCGTGAAGCCTTTGAAATATGATACCTGGCTTAC 1140
DB 340 GluGluGluIleMetLysLysIleArgGluSerTrpGluAsnAspIleAlaSerMetAsn 359
QY 1141 ACCCACTCC-----GGATACCTTC----- 1158
DB 360 LeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProThr 379
QY 1159 -----TCCAGCCTGTACCCCATCACCAAGTTGGCCGCTTCCCGCAT 1200
DB 380 SerGlyMetProProProThrSerGlyProProSerAlaMetThrProProTrpProGln 399
QY 1201 CATCACTCTTATCCAGACGAGAGATGTGTGATCTCTTATCCCAACCAAGGCTGTGAGC 1260
DB 400 PheGluGln---SerGluThrGluThrValHisLeuPheIleProAlaLeuSerValGly 418
QY 1261 GCCATCATCGGGAGAGAGGCGGACACATCAACACAGCTGGCAGATTCGCGGAGCTCT 1320
DB 419 AlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgPheAlaLysSer 438
QY 1321 ATCAAGATTTGCCCTCGGAGAGCCCGACAGCTGACGAAAGAGATGTCTATCAACCGGG 1380
DB 439 IleLysIleAlaProAlaGluValProAspAlaLysValArgMetValIleIleThrGly 458
QY 1381 CCACCGAAGCCCACTTCAAGGCCCGACGAGACGATCTTTGGGAAACTGAAAGAGAAAC 1440
DB 459 ProProGluAlaGlnPheLysAlaGlnGlyArgIleTrpGlyLysIleLysGlnGluAsn 478
QY 1441 TTCTTTAACCCCAAGAAAGATGAAAGCTGAAAGCCATATCATGAGTGCCTTCCACA 1500
DB 479 PheValSerProLysGluValLysLeuGlnAlaHisIleArgValProSerPheAla 498
QY 1501 GCTGGCCGGGTATTTGGCAAGAGGTGCAAGACCTGTAAAGCAATTCAGAACTTAACACT 1560
DB 499 AlaGlyArgValIleGlyLysGlyLysThrValAsnGluLeuGlnAsnLeuSerSer 518
QY 1561 GCAGAAATCATGCTGCTCGTACCAAGCCGATGAAAGAAATGAGGAAGATCGTCAGA 1620
DB 519 AlaGluValValProArgAspGlnThrProAspGluAsnAspGlnValValLys 538
QY 1621 ATTATCGGGCACTTCTTGTAGCCAGACTGCAAGCGCAAGATCAAGGAAATTTGACAA 1680
DB 539 IleThrGlyHisPheTrpAlaCysGlnValAlaGlnArgLysIleGlnGluIleLeuThr 558
QY 1681 CAGGTGAAGCAGACAGACAGAAA 1704
DB 559 GlnValLysGlnHisGlnGlnGln 566

```

RESULT 15

ADA28536 standard; protein; 579 AA.

ADA28536:

20-NOV-2003 (first entry)

Recombinant human lung tumour protein U523S #1.

Cancer; lung cancer; gene therapy; vaccine; human;

lung squamous cell carcinoma.

OS Homo sapiens.
XX
XX US2003064947-A1.
XX
XX 03-APR-2003.
PD
XX
XX 30-NOV-2001; 2001US-00007700.
XX
XX 18-MAR-1998; 98US-00040802.
PR 27-JUL-1998; 98US-00123912.
PR 22-DEC-1998; 98US-00221107.
PR 02-APR-1999; 99US-00285479.
PR 17-DEC-1999; 99US-0046396.
PR 30-DEC-1999; 99US-00476496.
PR 10-JAN-2000; 2000US-00480884.
PR 22-FEB-2000; 2000US-00510376.
PR 04-APR-2000; 2000US-00542615.
PR 28-JUN-2000; 2000US-00606421.
PR 02-AUG-2000; 2000US-00630940.
PR 21-AUG-2000; 2000US-00643597.
PR 15-SEP-2000; 2000US-00662786.
PR 09-OCT-2000; 2000US-00685696.
PR 12-DEC-2000; 2000US-00735705.
PR 07-MAY-2001; 2001US-00850716.
PR 28-JUN-2001; 2001US-00897778.
XX
XX (CORI-) CORIXA CORP.
XX
XX Wang T, Wang A, Skeiky YAM, Li SX, Kalos MD, Henderson RA,
PI McNeill PD, Fanger N, Retter MW, Durham M, Fanger GR, Vedvick TS;
PI Carter D, Matanabe Y, Peckham DW, Cai F, Foy TM;
XX
XX WPI; 2003-540798/51.
DR
XX New isolated polynucleotides and polypeptides useful for diagnosing,
PT preventing and/or treating cancer, particularly lung cancer.
XX
XX
XX Claim 9; Page 285-287; 296pp; English.
XX
XX The invention describes isolated polynucleotides and polypeptides useful
CC for diagnosing, preventing and/or treating cancer, particularly lung
CC cancer. A new isolated polynucleotide comprises: any of the 22 fully
CC defined nucleotide sequences (e.g. 1012, 900 or 2773 bp) given in the
CC specification; complements of the nucleotide sequences cited above; at
CC least 10 contiguous residues of the nucleotide sequences cited above; at
CC sequence that hybridise to any of the nucleotide sequences cited above; a
CC stringent conditions; a sequence that is at least 75 or 90% identical to
CC the above nucleotide sequences; or degenerate variants of the above
CC nucleotide sequences. The composition and methods are useful in
CC diagnosing, preventing and/or treating cancer, particularly lung cancer,
CC in gene therapy and in vaccines. This is the amino acid sequence of a
CC recombinant human lung tumour associated protein.
XX
XX
SQ Sequence 579 AA;
Alignment Scores:
Pred. No.: 5 47e-165
Score: 1875.00 Length: 579
Percent Similarity: 78.70% Matches: 369
Best Local Similarity: 64.96% Conservative: 78
Query Match: 31.86% Mismatches: 95
DB: 7 Indels: 26
Gaps: 6
US-09-270-437D-8 (1-3283) x ADA28536 (1-579)
QY 73 ATGAACAAGCTTTACATCGGGAACCTGAGCCCGCGTCACCGCCGACGACTCCGGCAG 132
DB 1 MetAsmyleutrylleglyAsnleuSerGlnAsnAlaIalProSerzAspleuGlnSer 20
QY 133 CTCCTTGGGAGACAGGAAGCTGCCCGCGGACAGGCTCTGCTGAAGTCCGGCTACGCC 192
DB 21 IlePheIysAspAlaIylsIleProValSerGlyProPheIeuValIylsThrGlyTyrAla 40

QY 193 TTCGTGACTACCCCGACAGAACTGGGCGCATCCGGCCATCGAACCCCTCGGGTAA 252
DB 41 PheValAspCysProAspGluSerTrpAlaIleGlyAlaIleuSerGlyIys 60
QY 253 GTGAATTGACGGGAAATATCGAAGTTGATTACTAGTCTCTAAAGAGGACC 312
DB 61 IleGluLeuHsIsglyLysProIleGluValGlnHsSerValProIylsArgGlnArgIle 80
QY 313 AGGAAATTCAGATTTCGAATCCATCCCTCCTCAGCTGAGGAGGTGTGATGACATT 372
DB 81 ArgLysIeuGlnIleArgAsnIleProProHsIleuGlnITPGLIValIleuAspSerIeu 100
QY 373 TTGGCTCAATATGGACAGTGGAGATGTGAAACAAGTCAACACAGACACAGAAACCGCC 432
DB 101 IeuValGlnTyrGlyValValGlnSerCysGlnGlnValAsnThrAspSerGlnThrAla 120
QY 433 GTTGTCACAGTCACATATGCAACAGAAAGAAAGCAAAATATGCCATGGACAAAGCTAAC 492
DB 121 ValValAsnValThrTyrSerSerIysAspGlnAlaArgGlnAlaIleuAspIylsIeuAsn 140
QY 493 GGGCATCAGTTTGAAGAACTACTCTTCAAGATTTCTCATCCGCGATGAGAGAGGTGAGC 552
DB 141 GlyPheGlnIleuGlnAsnPheThrIleuIylsValAlaTyrIleProAspGlnThrAlaAla 160
QY 553 TCCCTTGGCCCTCAG-----CGAGCCAGCGT-----GGGACCACTCTTCCCGG 600
DB 161 GlnGlnAsnProIeuGlnGlnProArgGlyArgArgGlyIleuGlnArgGlySerSer 180
QY 601 GAGCAAGGCGACGCCCCCTGGGGGCACTTCAGAGCGACACAGATTGATTCCTCGTGGG 660
DB 181 ArgGlnGly---SerProGlySerValSerIysGlnIylsProCysAspIeuProIeuArg 199
QY 661 ATCTGTGTCACCCACCGAGTTTGTGTGCATCATCGGAAGAGAGGCGTGCATTAAG 720
DB 200 IeuIeuValProThrGlnPheValGlyAlaIleIleGlyIylsGlnGlyAlaThrIleArg 219
QY 721 AACATCATAGCAGACCCAGCTCCCGGATAGATATCATAGAAAGAAAGAACTCGAGCT 780
DB 220 AsnIleThrIylsGlnThrGlnSerIylsIleAspValHsArgIylsGlnAsnAlaGlyAla 239
QY 781 GCAGAGAAAGCTGTCACCATTCATGCGCAACCCCAAGGGGACTTTGAAGCATGCGCGATG 840
DB 240 AlaGlnIylsSerIleThrIleIleuSerThrProIuGlnIylsThrSerAlaIylsSer 259
QY 841 ATTCTTGAATTCATGCAAGAAAGAGGAGATGACCAAACTAGCCGGAAGATTCCTGTG 900
DB 260 IleIeuIuIleuIleuHsIylsGlnAlaGlnAspIleIylsPheThrGlnIuIleProIeu 279
QY 901 AAAATCTGGCACAACAATGCTGTGTGGAAGACTGATTGGAAAAGAGCAGAAATTTG 960
DB 280 LysIleIeuAlaHsIAsnAsnPheValGlyArgIeuIleGlyIylsGlnGlyArgAsnIeu 299
QY 961 AAGAAATTCGAACATGAACAGAGGACCAAGTAACATTCATCTTTCGAGATTGAGC 1020
DB 300 LysLysIleGlnGlnAspThrAspThrLysIleThrIleSerProIeuGlnGlnIleuThr 319
QY 1021 ATATACAAACCCGGAAGAAACCATCATCTGTAAGGGGACAGTTGAGGCTTGCGCAGGCT 1080
DB 320 IeuTyrAsnProGlnIuHsThrIleThrValIylsGlnAsnValGlnIuThrCysAlaIylsAla 339
QY 1081 GAGATTCAGATTATGAAGAAAGCTGCGTAGAGCCTTTGAAAATGATATGATGCTGTATAC 1140
DB 340 GlnGlnGlnIuIleuIleuIylsIleArgIylsSerGlnIuIleuAsnIleAlaSerMetAsn 359
QY 1141 ACCCATCC-----GATATCTC----- 1158
DB 360 IeuGlnAlaHsIleuIleProGlyIleuAsnIleuAsnAlaIeuGlyIeuPheProThr 379
QY 1159 -----TCAGGCTGTACCCCGCATCACAGATTGAGCGCTGCGCGCAT 1200
DB 380 SerGlyMetProProIuHsThrIleThrIylsProProSerAlaMetThrProIylsProGln 399
QY 1201 CATCACTCTTATCCAGAGGAGGAGATTGTGAATCTCTCATCCCAACCAAGGCTGTGGCG 1260

```
Db      400 PheGluGln---SergIuThrGluThrValHisLeuPheIleProAlaLeuSerValGly 418
QY      1261 GCCATCATCGGGAGAAAGGGGACACATCAACAGCTGGCGAGATTGCCGGAGCTCT 1320
Db      419 AlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSer 438
QY      1321 ATCAAGATTGCCCTCGCGAAGGGCCAGACGTCAGCGAAAGATGGTCATCATCAACCGGG 1380
Db      439 IleLysIleAlaProAlaGlnAlaProAspAlaLysValArgMetValIleIleThrGly 458
QY      1381 CCACCGGAGAGCCCGAGTTCAAGGCGCCAGATCTTTGGGAAACTGAAAGGAAAG 1440
Db      459 ProProGlnAlaGlnPheLysAlaGlnGlyArgIleTyrglyLysIleLysGlnGln 478
QY      1441 TTCTTTAACCCCAAGAAAGAGTGAAGCTGAGAGCCATATCAGAGTGCCCTCTTCACA 1500
Db      479 PheValSerProLysGlnGlnValLysLeuGlnAlaHisIleArgValProSerPheAla 498
QY      1501 GCTGCGCGGGTGATTGGCAAGAGTGGCAAGACCGTGAACGAATCGCAAGACTTAACGAGT 1560
Db      499 AlaGlyArgValIleGlyLysGlyLysThrValAsnGlnLeuGlnAsnLeuSerSer 518
QY      1561 GCGAAGATCATGCTGCTGCTGACCAAGCCAGATGAAATAGAGAAAGTGATCGTCAGA 1620
Db      519 AlaGlnValValValProArgAspGlnThrProAspGlnAsnAspGlnValValLys 538
QY      1621 ATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGAAATTGTACAA 1680
Db      539 IleThrGlyHisPheTyrlaCysGlnValAlaGlnArgLysIleGlnGlnIleLeuThr 558
QY      1681 CAGGTGAGCAGCAGGACAGAAA 1704
Db      559 GlnValLysGlnHisGlnGlnGln 566
```

Search completed: July 23, 2004, 11:05:23
Job time : 180.09 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - protein search, using frame_plus n2p model

Run on: July 23, 2004, 10:56:36 ; Search time 29.5087 seconds
(without alignments)

11487.328 Million cell updates/sec

Title: US-09-270-437D-8

Perfect score: 5886

Sequence: 1 ggcagcgagagagcgagga.....aaccttgaatgttattt 3283

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cg2_1/USPRO.spool_p/US09270437/runat_23072004.095740.27466/app_query.fasta_1.5582
-DB=Issued_Patents_AA -QWMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPTH=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=Pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US09270437 @CGN 1.1 59 @runat_23072004.095740.27466 -NCPD=6 -ICPD=3
-NO MMAP -LARGEOUTRY -NEG_SCORES=0 -WAIT -DSFBLCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA:*
1: /cg2_6/ptodata/2/iaa/5A.COMB.rep:*
2: /cg2_6/ptodata/2/iaa/5B.COMB.rep:*
3: /cg2_6/ptodata/2/iaa/6A.COMB.rep:*
4: /cg2_6/ptodata/2/iaa/6B.COMB.rep:*
5: /cg2_6/ptodata/2/iaa/PCTUS.COMB.rep:*
6: /cg2_6/ptodata/2/iaa/backfile1.rep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1934	32.9	577	3	US-09-261-855-2
2	1875	31.9	579	4	US-09-643-597-348
3	1875	31.9	579	4	US-09-542-615A-348
4	1875	31.9	579	4	US-09-606-421B-348
5	1868	31.7	579	4	US-09-643-597-176
6	1868	31.7	579	4	US-09-480-884A-176
7	1868	31.7	579	4	US-09-542-615A-176
8	1868	31.7	579	4	US-09-606-421B-176
9	263.5	4.5	644	1	US-08-021-608D-2
10	263.5	4.5	644	1	US-08-726-160-2
11	263.5	4.5	644	5	PCT-US94-01782-2
12	262	4.5	643	1	US-08-021-608D-10

13	262	4.5	643	1	US-08-726-160-10	Sequence 10, Appl
14	262	4.5	643	5	PCT-US94-01782-10	Sequence 10, Appl
15	255	4.3	590	1	US-08-021-608D-8	Sequence 8, Appl
16	255	4.3	590	1	US-08-726-160-8	Sequence 8, Appl
17	255	4.3	590	5	PCT-US94-01782-8	Sequence 8, Appl
18	215	3.7	48	3	US-09-261-855-20	Sequence 20, Appl
19	210	3.6	48	3	US-09-261-855-24	Sequence 24, Appl
20	210	3.6	48	3	US-09-261-855-18	Sequence 18, Appl
21	207	3.5	530	1	US-08-187-793-4	Sequence 4, Appl
22	201	3.4	47	3	US-09-261-855-21	Sequence 21, Appl
23	200	3.4	49	3	US-09-261-855-22	Sequence 22, Appl
24	197	3.3	47	3	US-09-261-855-17	Sequence 17, Appl
25	194.5	3.3	720	4	US-09-252-991A-21881	Sequence 21881, A
26	190.5	3.2	705	4	US-09-252-991A-30792	Sequence 30792, A
27	179	3.0	697	4	US-09-252-991A-24009	Sequence 24009, A
28	178.5	3.0	657	4	US-09-252-991A-28001	Sequence 28001, A
29	177	3.0	47	3	US-09-261-855-23	Sequence 23, Appl
30	176	3.0	47	3	US-09-261-855-19	Sequence 19, Appl
31	176	3.0	742	4	US-09-252-991A-32659	Sequence 32659, A
32	173.5	2.9	1706	4	US-09-252-991A-31760	Sequence 31760, A
33	170.5	2.9	343	1	US-08-187-793-2	Sequence 2, Appl
34	169	2.9	1476	4	US-09-252-991A-29427	Sequence 29427, A
35	168.5	2.9	745	4	US-09-252-991A-29790	Sequence 29790, A
36	168.5	2.9	748	4	US-09-252-991A-18427	Sequence 18427, A
37	167.5	2.8	1228	4	US-09-252-991A-17764	Sequence 17764, A
38	167	2.8	987	4	US-09-252-991A-27403	Sequence 27403, A
39	166	2.8	783	4	US-09-252-991A-18035	Sequence 18035, A
40	166	2.8	871	4	US-09-252-991A-19431	Sequence 19431, A
41	164	2.8	654	4	US-09-252-991A-25423	Sequence 25423, A
42	164	2.8	830	4	US-09-252-991A-27142	Sequence 27142, A
43	163	2.8	685	4	US-09-252-991A-31814	Sequence 31814, A
44	162.5	2.8	955	4	US-09-252-991A-24254	Sequence 24254, A
45	162	2.8	681	4	US-09-252-991A-24567	Sequence 24567, A

ALIGNMENTS

RESULT 1
US-09-261-855-2
Sequence 2, Application US/09261855A
Patent No. 6255055
GENERAL INFORMATION:
APPLICANT: ROSS, Jeffrey
TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN
FILE OF INVENTION: (CRD-BP) AND ITS NUCLEIC ACID SEQUENCE
FILE REFERENCE: 960296.95131
CURRENT APPLICATION NUMBER: US/09/261.855A
CURRENT FILING DATE: 1999-03-03
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 577
TYPE: PRT
ORGANISM: Mus musculus
US-09-261-855-2

Alignment Scores:

Pred. No.: 9.92e-163
Score: 1934.00
Percent Similarity: 79.10%
Best local Similarity: 65.80%
Query Match: 32.86%
DB: 3
Length: 577
Matches: 381
Conservative: 77
Mismatches: 95
Indels: 26
Gaps: 5

US-09-270-437D-8 (1-3283) x US-09-261-855-2 (1-577)

QY 73 ATGAAGAGCTTACATCGGAACCTGAGCCCGCTGACCGGACGACCTCCGCGAG 132
DB 1 MetAsnIlyLeuYrIleGlyAsnIeuAsnGuservValThrProAlaSpLeuGIntys 20
QY 133 CTCCTTGGGAGCAGGAGCTGCCCTCGGGGAGCAGACTCTGCTGAAGTCCGCTACGCC 192
DB 21 ValpheAlaGlnHisIlyIleSerYrIserGlyGlnpheIeuValIysSerGlyTyrAla 40

QY 193 TTGGTGAAGTACCCCGACAGAACTGGCCATCCGCCATCGAGACCTCTCGGGTAA 252
 Db PheValAspCysProAspGluHisTyrPheValMetLysAlaIleGluThrPheSerGlyLys 60
 QY 253 GTGGAATTGCAATGGGAAATATGATGGAAGTATTAATCACTGCTTAAAGCTAAAGAGC 312
 Db ValGluLeuGlnGlyLysArgLeuGlnMetGlnHisSerValProLysLysGlnArgSer 80
 QY 313 AGGAAATTCAGATTCGAACATCCCTCTCACTCGATGGAGAGTGTGGATGACTT 372
 Db ArgLysIleGlnIleArgAsnIleProProGlnLeuArgTyrGlnValLeuAspSerLeu 100
 QY 373 TTGGCTCAATATGGAACAGTGAAGATGTGAACAGTCAACAGACAGACAGAAACCCGC 432
 Db LeuAlaGlnTyrGlyThrValGlnAsnCysGlnGlnValAsnThrGlnSerGlnThrAla 120
 QY 433 GTTGTCAAGTCACATATGCAACAGAGAGAGAAATAGCATGAGAGAGTAAAC 492
 Db ValValAsnValThrTyrSerAsnArgGlnGlnThrArgGlnAlaIleMetLysLeuAsn 140
 QY 493 GGGCATCATGTTGAGAACTACTCTTCAAGATTTCTCATCTCCGATGAGAGAGTGAAC 552
 Db GlnHisGlnLeuGlnHisAlaLeuLysValSerTyrIleProAspGlnGlnIleThr 160
 QY 553 TCCCTCTGCGCCCTCAGCGAGCCAGCGTGGGAGCACTCTCCCGGAGAGAGGCCAC 612
 Db 161 -----GlnGlyProGlnAsnGlyArgArgGlyGlyPheGlySerArgGlyGlnProArg 178
 QY 613 -----GCCCTGGGGGCACTCTCAAGGAGAGAGAGAGAGATTTCCCGCTG 657
 Db 179 GlnGlySerProValAlaAlaGlyAlaProAlaLysGlnGlnProValAspIleProLeu 198
 QY 658 CGGATCTGATGATCCCGACCCAGATTTGTGTGTCATCATCGAAAGAGAGGCTTGACCA 717
 Db 199 ArgLeuLeuValProThrGlnTyrValGlyAlaIleIleGlySerGlnGlyAlaThrIle 218
 QY 718 AAGACATCATCTAAGACAGACCCAGTCCCGGTGATATCCATGAAAGAGAGAGCTGGA 777
 Db 219 ArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgLysGlnAsnAlaGly 238
 QY 778 GGTGAGAGAGAGCGTGTGACATCATGACACCCCGAGAGAGGAGCTTCTGAGAGTCCG 837
 Db 838 ATGATCTTGAATCATGACAGAAAGAGAGAGATGAACCAACTAGCCGAGAGATTTCT 897
 Db 259 MetIleLeuGlnIleMetHisLysGlnAlaLysAspThrLysThrAlaAspGlnValPro 278
 QY 898 GTGAAATCTTGGCACACATGGCTGTGTAAGAGTGAATGGAAAGAGAGAGAGAT 957
 Db 279 LeuLysIleLeuAlaHisAsnAspPheValGlyArgLeuIleGlyLysGlnGlyArgAsn 298
 QY 958 TTGAGAGAAATGTAACATGAACAGGAGACCAAGTAACTCATCTTGTGACAGATTG 1017
 Db 299 LeuLysLysValGlnGlnAspThrGlnThrLysIleThrIleSerSerLeuGlnAspLeu 318
 QY 1018 AGCATATCAACCCCGAAAGAACATCATCTGTGAAGGAGACATGAGGCTGTGCCAGT 1077
 Db 319 ThrLeuLysAsnProGlnArgThrIleThrValLysGlyAlaIleGlnAsnCysCysArg 338
 QY 1078 GCTGAGATTAAGATTATGAAGAGAGCTGCGTGAAGGCTTTGAAATGATATGAGCTGTT 1137
 Db 339 AlaGlnGlnGlnIleMetLysValArgGlnAlaTyrGlnAsnAspValAlaAlaMet 358
 QY 1138 AAGACCACTCC-----GGATCTTC----- 1158
 Db 359 SerLeuGlnSerHisLeuIleProGlyLeuAsnLeuAlaValGlyLeuPheProAla 378
 QY 1159 ---TCCAGCTGTACCCCGATCACCAG-----TTGGCCCGTTCGCGAT 1200
 Db 379 SerSerSerAlaValProProProSerSerSerValThrGlyAlaAlaProTyrSerSer 398

QY 1201 CATCACTCTTATCCAGACAGAGAGATTGTGAATCTTCAATCCCAACCCAGAGCTGTGGGC 1260
 Db 399 PheMetGlnAlaProGlnGlnGlnMetValGlnValPheIleProAlaGlnAlaValGly 418
 QY 1261 GCCATCATCGGAGAGAGGGGCGACATCAAAACAGCTGGGAGATTGCGCGAGCTCT 1320
 Db 419 AlaIleIleGlyLysGlyGlnHisIleLysGlnLeuSerArgPheAlaSerAlaSer 438
 QY 1321 ATCAAGATTGCGCTGCGGAGAGGCCAGACGTCAGCGGAAGAGATGTCAATCATCCGCG 1380
 Db 439 IleLysIleAlaProProGlnThrProAspSerLysValArgMetValValIleThrGly 458
 QY 1381 CCACCGAAGCCCACTTCAAGGCCAGGAGCGATCTTTGGGAACTGAAAGAGAGAAAC 1440
 Db 459 ProProGlnAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysLeuLysGlnGln 478
 QY 1441 TTCTTAAACCCCAAGAGAGTGAAGCTGGAAGCGCATATCAAGATGCGCTCTTCCACA 1500
 Db 479 PhePheGlyProLysGlnGlnValLysLeuGlnThrHisIleArgValProAlaSerAla 498
 QY 1501 GCTGCGCGGATGATGGCAAGAGTGGCAAGACCGTGAACGAACCTGACAGACTTAACAGT 1560
 Db 499 AlaGlyArgValIleGlyLysGlyGlyLysThrValAsnGlnLeuGlnAsnLeuThrAla 518
 QY 1561 GCGAAGTCAATGTCCTCTGTCGACCAACCGCATGAAATGAGAGAGTGAATCGTCAGA 1620
 Db 519 AlaGlnValValValProArgAspGlnThrProAspGlnAsnAspGlnValIleValLys 538
 QY 1621 ATTATCGGGCACTCTGTTGTCGACAGACTGACACCGCAAGTCAAGAGAAATTGTACAA 1680
 Db 539 IleIleGlyHisPheTyrAlaSerGlnMetAlaGlnArgLysIleArgAspIleLeuAla 558
 QY 1681 CAGGTGAAGCAGCAGAGAGCAAAATACCTCAGGAGTCCCTCACAGGCGACGAAG 1737
 Db 559 GlnValLysGlnGlnHisGlnLysGlyGlnSerAsnLeuAlaGlnAlaArgArgLys 577

RESULT 2
 US-09-643-597-348
 ; Sequence 348, Application US/09643597
 ; Patent No. 6426072
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Tongtong
 ; APPLICANT: Fan, Ligu
 ; APPLICANT: Kalos, Michael D.
 ; APPLICANT: Bangur, Chaitanya S.
 ; APPLICANT: Hosken, Nancy
 ; APPLICANT: Fanger, Gary R.
 ; APPLICANT: Li, Samuel X.
 ; APPLICANT: Wang, Aijun
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Henderson, Robert A.
 ; APPLICANT: McNeill, Patricia D.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
 ; FILE REFERENCE: 210121.455C11
 ; CURRENT APPLICATION NUMBER: US/09/643,597
 ; NUMBER OF SEQ ID NOS: 369
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 348
 ; LENGTH: 579
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-643-597-348

Alignment Scores: 1.7e-157 Length: 579
 Pred. No.: 1875.00 Matches: 369
 Score: 1875.00 Matches: 369
 Percent Similarity: 78.70% Conservative: 78
 Best Local Similarity: 64.96% Mismatches: 95
 Query Match: 31.86% Indels: 26
 DB: 4 Gaps: 6

US-09-270-437D-8 (1-3283) x US-09-643-597-348 (1-579)

```

QY 73 ATGAAGAGCTTTTACATCGGGAACCTGAGCCCGCGCTCACCGCCAGCAAGCTCCGGCAG 132
Db 1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaIleProSerHisPheLeuGlySer 20
QY 133 CTCTTTGGGGACAGGAAGCTGCCCTGGCGGACAGGTCCTGCTGAAAGTCGCGCTACGCC 192
Db 21 IlePheLysAspAlaLysIleProValSerGlyProPheLeuValLysThrGlyTyrAla 40
QY 193 TTGCGTGAATACCCCGACGAGAACTGGCCATTCGCGCCATCGAGACCTCTCGGGTAA 252
Db 41 PheValAspCysProAspGlnSerIlePheValLeuLysAlaIleGluAlaLeuSerGlyLys 60
QY 253 GTGGAATTGCATGGGAAAATCATGGAAGTTGATTATTCAGTCTCTTAAAAAGCTAGAGAC 312
Db 61 IleGluLeuHisGlyLysProIleGluValGluHisSerValProLysAspGlnThrGly 80
QY 313 AGGAAAATTCAGATTTCGAAAATCCCTCTCTACCTCGACGTGGAGAGGTGTGGATGACTT 372
Db 81 ArgLysLeuGlnIleArgAsnIleProPheHisLeuGlnIleProGluValLeuAspSerLeu 100
QY 373 TTGGCTCATATATGGACAGTGGAGAAATGTGGAACAAGTAAACAACAACAAGACCCGCC 432
Db 101 LeuValGlnTyrGlyValValGlnSerCysGlnGlnValAsnThrAspSerGlnThrAla 120
QY 433 GTTGTCAACGTCACTATATGCAACAAGAGAAAGCAAAATATGCCATGGAGAGGCTAAC 492
Db 121 ValValAsnValThrTyrSerSerLysAspGlnAlaArgGlnAlaLeuAspLysLeuAsn 140
QY 493 GGGCATCATGTTGAGAACTACTCTCTTCAAGATTTCCTCATCTCCGATGAGAGGTGAGC 552
Db 141 GlyPheGlnLeuGlnAsnPheThrLeuLysValAlaTyrIleProAspGlnThrAlaAla 160
QY 553 TCCCTTGCGCCCGCTCAG-----CGAGCCAGCGT-----GGGACCACTCTTCCGG 600
Db 161 GlnGlnAsnProLeuGlnGlnProArgGlyArgArgGlyLeuGlnArgGlySerSer 180
QY 601 GAGCAAGGCCACGCCCCCTGGGGCACTTTCAGGCGCAAGCATTTGATTCCCGCTGGC 660
Db 181 ArgGlnGly--SerProGlySerValSerLysGlnLysProCysAspLeuProLeuArg 199
QY 661 ATCTGTGTCCTCCACCCAGTTTGTGTGTCATCATCGGAAGAAGGCGCTTGACCATTAAG 720
Db 200 LeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGlnGlyAlaThrIleArg 219
QY 721 AACATCTAGACAGACCCAGTCCCGGGTACATATCCATAGAAAAGAACTGTGAGCT 780
Db 220 AsnIleThrLysGlnThrGlnSerLysIleAspValHisArgLysGlnAsnAlaGlyAla 239
QY 781 GCAGAGAAGCCTGTACCATCATCCAGCCCAAGGGGACTTGTGAAGCATGCCGATG 840
Db 240 AlaGlnLysSerIleThrIleLeuSerThrProGlnGlyThrSerAlaAlaCysLysSer 259
QY 841 ATTCTTGAATCATGCAAGAAAGAGCATGAGACCAACATAGCCGAGAGATTCCCTGTG 900
Db 260 IleLeuGlnIleLeuMetHisLysGlnAlaGlnAspLysPheThrGlnGlnIleProLeu 279
QY 901 AAAATCTGGCACACAATGCGCTGTGGGAAGCTGATTGGAAGAAAGAGCAAGAAATTG 960
Db 280 LysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGlnGlyArgAsnLeu 299
QY 961 AAGAAAATTGAAATGAAACAGGAGCAACAGATTAACAATCATCTTTGGAGATTGGAGC 1020
Db 300 LysLysIleIleGlnGlnAspThrAspThrLysIleThrIleSerProLeuGlnGlnLeuThr 319
QY 1021 ATTTAACAACCCGAGAAAGAACATCACTGTGAAAGGACAGATTGAGGCTGTGCAGTGT 1080
Db 320 LeuTyrAsnProGlnArgThrIleThrValLysGlyAsnValGlnThrCysAlaLysAla 339
QY 1081 GAGATGAGATTATGAAAGAGCTGCGTGGAGCCTTTGAAATATATATCTGCGCTGTTAAC 1140
Db 340 GlnGlnGlnIleLeuTyrLysIleArgGlnSerTyrGlnAsnAspIleAlaSerMetAsn 359

```

```

QY 1141 ACCCACTCC-----CGATACCTTC----- 1158
Db 360 LeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProThr 379
QY 1159 -----TCCAGCGGTACCCCCCATACCAAGTTTGCCCGTCCGGCAT 1200
Db 380 SerGlyMetProProProThrSerGlyProProSerAlaMetThrProProTyrProGln 399
QY 1201 CATCACTCTTATTCAGAGCAGAGATTGAAATCTTCAATCCCAACCCAGGCTGTGGCC 1260
Db 400 PheGlnGln--SerGlnThrGlnThrValHisLeuPheIleProAlaLeuSerValGly 418
QY 1261 GCCATCATCGGAGAGAGAGGAGCAACATCAACAGCTGGCGAGATTCCCGAGGCTCT 1320
Db 419 AlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSer 438
QY 1321 ATCAAGATTGGCCCGCGGAAAGCCCAAGCTGCAGGAAAAGATGCTCATCATCCGGG 1380
Db 439 IleLysIleAlaProAlaGlnAlaProAspAlaLysValArgMetValIleIleThrGly 458
QY 1381 CCACCGAAGCCCAAGTTCAGAGCCCGAGGAGCATCTTTGGAGAACTGAAAGAGAAAC 1440
Db 459 ProProGlnValGlnPheLysAlaGlnGlyArgGlyTyrGlyLysIleLysGlnGlnAsn 478
QY 1441 TTCTTTAAACCCCAAGAAAGATGAAAGCTGGAAGGCAATTCAGAGTCCCTCTTCACA 1500
Db 479 PheValSerProLysGlnGlnValLysLeuGlnAlaHisIleArgValProSerPheAla 498
QY 1501 GCTGGCGGGGTGATTGGGAAAGCTGGCAAGACCGTGAACGAACCTGCAGAACTTAACAGT 1560
Db 499 AlaGlyAlaGValIleGlyLysGlyGlyLysThrValaGlnGlnLeuGlnAsnLeuSerSer 518
QY 1561 GCAGAAGTCATCGTGCCTCGTGACCAACGCGCAGATGAAATGAGAAAGTATCGTCAGA 1620
Db 519 AlaGlnValValValProArgAspGlnThrProAspGlnAsnAspGlnValValLys 538
QY 1621 ATTAATCGGGCACTTCTTTGTAGCCAGACTGCAGAGCGCAAGATCAGGAATTTGACAA 1680
Db 539 IleThrGlnHisPheTyrAlaCysGlnValAlaGlnIleArgLysIleGlnGlnIleLeuThr 558
QY 1681 CAGGTGAAGCAGCAGAGCAAGAAA 1704
Db 559 GlnValLysGlnHisGlnGlnGln 566

RESULT 3
US-09-542-615A-348
? Sequence 348, Application US/09542615A
? Patent No. 6518256
? GENERAL INFORMATION:
? APPLICANT: Wang, Tongtong
? APPLICANT: Pan, Liqun
? APPLICANT: Kalos, Michael D.
? APPLICANT: Bangur, Chaitanya S.
? APPLICANT: Hosken, Nancy A.
? APPLICANT: Fanger, Gary R.
? TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
? FILE REFERENCE: 210121.455C8
? CURRENT APPLICATION NUMBER: US/09/542.615A
? CURRENT FILING DATE: 2000-04-14
? NUMBER OF SEQ ID NOS: 350
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 348
? LENGTH: 579
? TYPE: PRT
? ORGANISM: Homo sapiens
US-09-542-615A-348

Alignment Scores:
Pred. No.: 1,76-157 Length: 579
Score: 1875.00 Matches: 369
Percent Similarity: 78.708 Conservative: 78

```


Best Local Similarity: 64.96% Mismatches: 95
 Query Match: 31.86% Indels: 26
 DB: 4 Gaps: 6

US-09-270-437d-8 (1-3283) x US-09-542-615A-348 (1-579)

```

QY 73 ATGAACAAGCTTTACATTCGGAGAACCTTGAGCCCGCCGTCACCGCCGACGACCTCGGAGC 132
Db 1 MetAsnLysLeuTyrTleGlyAsnLeuSerGluAsnAlaAlaProSerAspLeuGluSer 20
QY 133 CTCCTTGGGACAGAGAACTGCCCCCTGGGAGAGAGCTCTGCTGAAGTCCGGCTAGCC 192
Db 21 IlePheLysAspAlaLysIleProValSerGlyProPheLeuValLysThrGlyTyrAla 40
QY 193 TTCGTGACTACCCGACCGAATCTGGCCATCGCCGCTTCGAGACCCCTCCGGGTAA 252
Db 41 PheValAspCysProAspGluSerTyrPalaLeuLysAlaIleGluAlaLeuSerGlyLys 60
QY 253 GTGAATTCATGGGAAATATCATGAGTTGATTCAGTCTCTAAAGCTAAGGAGC 312
Db 61 IleGluLeuHsGlyLysProIleGluValGluHsSerValProLysArgGlnArgIle 80
QY 313 AGGAATTCAGATTGCAACAATCCCTCCTCACTGCACTGGAGCGTGTGATGAGCTT 372
Db 81 ArgLysLeuGlnIleArgAsnIleProProHsIleuGlnTyrGluValLeuAspSerLeu 100
QY 373 TTGGCTCAATATGGACAGTGAAGATGTGAAACAGTCAACAGACAGACAGAAACCGCC 432
Db 101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnHsAspSerGlnHsAla 120
QY 433 GTTGTCACATGCACTATGCAACAAGAGAAAGAAATAGCCATGAGAGAGCTAAC 492
Db 121 ValValAsnValThrTyrSerSerLysAspGlnAlaArgGlnAlaLeuAspLysLeuAsn 140
QY 493 GGCGATCAGTTTGAAGACTACTCCTTCAAGATTCTCAATCCCGGATGAAGGCTGAGC 552
Db 141 GlyPheGlnLeuGlnAsnPheThrLeuLysValAlaTyrIleProAspGlnThrAlaAla 160
QY 553 TCCCTTCGCCCCCTGAG-----CGAGCCAGCGT-----GGGAGCACCTTCCCGG 600
Db 161 GlnGlnAsnProLeuGlnGlnProArgGlyArgArgGlyLeuGlyGlnArgLysSer 180
QY 601 GAGCAGCGCAGCCCGCCCTGGGCACTTCACAGCCAGACAGATGATTTCCCGCTGCGG 660
Db 181 ArgGlnGly--SerProGlySerValSerLysGlnLysProCysAspLeuProLeuArg 199
QY 661 ATCTGGTCCCAACCCAGTTTGTGTGTCATCATTCGAAAGAGAGGCTTGACCATTAAG 720
Db 200 LeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGlnGlyAlaThrIleArg 219
QY 721 AACATCACTAAGACAGCCAGTCCCGGCTAGATTCCATGAAAGAGAACTCGAGAGCT 780
Db 220 AsnIleThrLysGlnThrGlnSerLysIleAspValHsIleArgLysGluAsnAlaGlyAla 239
QY 781 GCGAGAGAGCGCTGCACATTCATCCAGCCAGGAGGAGCTTCTGAAGCATGCGGATG 840
Db 841 ATTCTGAAATCATGAGAAAGAGGAGCATGAGACCAACTAGCCGAGAGATCTCTG 900
Db 260 IleLeuGlnIleMetHsLysGluAlaGlnAspIleLysPheThrGlnGlnIleProLeu 279
QY 901 AAAATCTTGGACACATGCTTGGTGGAGACAGTATGGAAAGAGAGGAGAAATTTG 960
Db 280 LysIleLeuAlaHsAsnAsnPheValGlyArgLeuIleGlyLysGlnGlyArgAsnLeu 299
QY 961 AAGAAATGAACATGAGAGGAGCAAGATTAACATCTCATCTTTCAGATTGAGC 1020
Db 300 LysLysIleGlnGlnAspThrAspThrLysIleThrIleSerProLeuGlnGlnLeuThr 319
QY 1021 ATATACAAACCGGAAAGACATCATCTGTAAGAGGAGCATTTGAGGCTGTGCCAGTCT 1080
Db 320 LeuTyrAsnProGluArgThrIleThrValLysGlyAsnValGlnThrCysAlaLysAla 339

```

```

QY 1081 GAGATGAGATTATGAGAAAGCTGCGTGAAGGCTTTGAAAATGATATGCTGCTTTAAC 1140
Db 340 GlnGluGlnIleMetLysLysIleArgGluSerTyrGlnAsnAspIleAlaSerMetAsn 359
QY 1141 ACCCATCC-----GGATACCTC----- 1158
Db 360 LeuGlnAlaHsIleLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProProthr 379
QY 1159 -----TCCAGCCTGTACCCCATCAACAGTTTGGCCGCTCCCGCAT 1200
Db 380 SerGlyMetProProProThrSerGlyProProSerAlaMetThrProProLysProGln 399
QY 1201 CATCATCTTATCCAGACAGAGAGATTTGATCTCTTCATCCCAACCCAGCCTGGGC 1260
Db 400 PheGluGlnIleSerGlnThrValHsLeuPheIleProAlaLeuSerValGly 418
QY 1261 GCCATCATCGGAGAGAGAGGAGGAGCAATCAACAAACAGTGGGAGATTGCGCGGACCTT 1320
Db 419 AlalIleIleGlyLysGlnGlyGlnHsIleLysGlnLeuSerArgPheAlaGlyAlaSer 438
QY 1321 ATCAAGATTGCCCCGCGAGAGGCCAGACGTCAAGCAAGAGATGTCATCATCCCGG 1380
Db 439 IleLysIleAlaProAlaGlnAlaProAspAlaLysValArgMetValIleIleThrGly 458
QY 1381 CCACCGGAAACCCAGTTCAAGGCCCGAGCGGATCTTTGGGAAACTGAAGAGAAAC 1440
Db 459 ProProGlnAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLysGlnGluAsn 478
QY 1441 TTCCTTAACCCCAAGAGAGAGTGAAGCTGAAAGCGCATATCAGATGCTCTTCCACA 1500
Db 479 PheValSerProLysGlnGlnValLysLeuGlnAlaHsIleIleArgValProSerPheAla 498
QY 1501 GCTGCGCGGAGTATTTGGCAAGAGTGGCAAGCCGTGAAGAACTGACGAACTTAACAGT 1560
Db 499 AlaGlyArgValIleGlyLysGlyLysThrValAsnGlnLeuAsnLeuSerSer 518
QY 1561 GCAGAGTCACTGCTGCTGCTGACCAAGCCAGATGAAGATGAGAGATGATCTGACA 1620
Db 519 AlaGlnValValValProArgAspGlnThrProAspGlnAsnAspGlnValValLys 538
QY 1621 ATATATGGGCACTTTCTTGTAGCCAGACTGCAAGCGCAAGATCAAGGAAATTTGACA 1680
Db 539 IleThrGlyAsnPheTyrAlaCysGlnValAlaGlnArgLysIleGlnGlnIleLeuThr 558
QY 1681 CAGGTGAAGCAGCAGAGACAGAA 1704
Db 559 GlnValLysGlnHsGlnGlnGln 566

```

RESULT 4
 US-09-606-421B-348
 : Sequence 348, Application US/09606421B
 : Patent No. 6531315
 : GENERAL INFORMATION:
 : APPLICANT: Wang, Tonglong
 : APPLICANT: Fan, Liqun
 : APPLICANT: Kalos, Michael D.
 : APPLICANT: Bangur, Chaltanya S.
 : APPLICANT: Hosken, Nancy
 : APPLICANT: Fanger, Gary R.
 : APPLICANT: Li, Samuel X.
 : APPLICANT: Wang, Aijun
 : APPLICANT: Skeiky, Yasin A.W.
 : TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 : OF LUNG CANCER
 : FILE REFERENCE: 210121.455C9
 : CURRENT FILING DATE: US/09/606,421B
 : NUMBER OF SEQ ID NOS: 358
 : SOFTWARE: FastSeq for Windows Version 3.0
 : SEQ ID NO 348
 : LENGTH: 579
 : TYPE: PRT

ORGANISM: Homo sapiens
US-09-606-421B-348

Alignment Scores:
Pred. No.: 1,7e-157
Score: 1875.00
Percent Similarity: 78.70%
Best Local Similarity: 64.96%
Query Match: 31.86%
DB: 4
Length: 579
Matches: 369
Conservative: 78
Mismatches: 95
Indels: 26
Gaps: 6

US-09-270-437D-8 (1-3283) x US-09-606-421B-348 (1-579)

QY 73 ATGAACAAGCTTTAATCGGGAACCTGAGCCCGCGCTACCGCCGAGCACTCCGGGAC 132
Db 1 MetAnlysleuTylleGlyasnleuSerGluasnAlaAlaProSerAspneuGlnSer 20
QY 133 CTCTTTGGGGAGAGAAAGCTGCGCCCTGGCGGAGAGGTCCTGTAAGTCCGGCTAACGC 192
Db 21 llePheLysAspAlaLysleleProValSerGlyProPheleuValLysThrGlyTyrAla 40
QY 193 TTCCGGAAGTACCCCGACGAGAACTGGGGCAATCCCGCCGATGAGAACCTCTCGGGTAA 252
Db 41 PheValAspCysProAspGlnSerTyrPalaLeuLysAlaAlaGluAlaLeuSerGlyLys 60
QY 253 GTGAATTCAGTGGGAAATATCATGAAAGTGTATCTACGCTCTTAAAGCTAAGAGAGC 312
Db 61 lleGluLeuHisGlyLysProileGluValGlnHisSerValProLysArgGlnArgile 80
QY 313 AGGAAATTCAGATTCGAAACATCCCTCTCACTGACGAGGAGGTGGATGAGACTT 372
Db 81 ArgLysleuGlnleleArgAsnleleProProHisleuGlnTyrGluValleuAspSerleu 100
QY 373 TTGGCTCAATATGGGAGAGTGGAGATGTGAAACAAGTCAACAAGCAACAAGAACCGCC 432
Db 101 leuValGlnTyrGlyValValGlnSerCysGlnGlnValasnTyrAspSerGluThrAla 120
QY 433 GTTGTCAACGTCATATGCAACAAGAGAAAGCAAAATATAGCATGAGCAAGCTAAC 492
Db 121 ValValasnValThrTyrSerSerLysAspGlnAlaArgGlnAlaLeuAspLysleuasn 140
QY 493 GGGATCATGTTTGGAACTACTCTTCAAGATTTCTCAATCCCGATGAGAGAGTGAAGC 552
Db 141 GlyPheGlnleuGlnasnPheThrleuLysValAlaTyrleleProAspGlnThrAlaAla 160
QY 553 TCCCTTCGCGCCCTCGAG-----CGAGCCAGCGT-----GGGACCACTCTTCCCGG 600
Db 161 GlnGlnasnProleuGlnGlnProArgGlyArgArgleuGlyGlnArgGlySerSer 180
QY 601 GAGCAAGCGCAAGCCCTGGGGGCACTTCTCAGGCGCAGACAGATGATTTCCCGCTCGG 660
Db 181 ArgGlnGly---SerProGlySerValSerLysleuSerProCysAspLeuProleuArg 199
QY 661 ATCTGTGCTCCCAACCCGATTTGTTGTCATCATCGAAAGAGGAGCTTGAACATTAAG 720
Db 200 leuLeuValProThrGlnPheValGlyAlaIlelleLysGlnGluAlaThrIleArg 219
QY 721 AACATCACTAAGAGAACCCGATCCCGGGTATGATATCATGAAAGAACTCTGAGACT 780
Db 220 AsnIleThrLysGlnThrGlnSerLysIleAspValHisArgLysGlnasnAlaGlyAla 239
QY 781 GCAGAGAAGCTGTCAACCTCATGCAACCCGAGAGGAGCTTCTGAAGATCCCGCAG 840
Db 240 AlaGlnLysSerleleThrIleleuSerThrProGlnGlyThrSerAlaAlaCysLysSer 259
QY 841 ATTCTTGAATCATGCAAGAGAGGAGATGAGACCAAACTGAGCCGAGATTTCTCTG 900
Db 260 lleleuGlnlleleleHisLysGlnAlaGlnAspIleLysPheThrGlnGlnleleProleu 279
QY 901 AAAATCTTGCAACAATGCTTGTGTTGAAAGCTGATTTGAAAGAAAGAGGCAAAATTTG 960
Db 280 LysIleleuAlaHisasnasnPheValGlyArgleuIlelleLysGlnGlyArgasnleu 299

QY 961 AAGAAATTAATGACATGAAACAGGACCAAGATPACATCTTTGACGATTTGAGC 1020
Db 300 LysLysIleleGlnGlnAspThrAspThrLysIleThrIleSerProleuGlnleuThr 319
QY 1021 ATATCAACCCGGGAAAGAACCTCATCTGTGAAGGGCAGACGTTGAGCCCTGTGCACTGCT 1080
Db 320 leuTyrAsnProGlnArgThrIleThrValLysGlyAsnValGlnThrCysAlaLysAla 339
QY 1081 GAGATAGAGATTTATGAAAGAGCTGCGTGAAGCTTGTGAAGATGATGCTGTAAAC 1140
Db 340 GlnGlnGlnlleleMetLysLysIleArgGlnSerTyrGlnasnAspIleAlaSerMetasn 359
QY 1141 ACCCACTCC-----GATACTTC----- 1158
Db 360 leuGlnAlaHisleuIleProGlyleuasnleuasnAlaLeuGlyleuPheProThr 379
QY 1159 -----TCCAGCTGTAAACCCCATCACAGATTTGGCCGTTCCCGCAT 1200
Db 380 SerGlyMetProProProThrSerGlyProProSerAlaMetThrProProTyrProGln 399
QY 1201 CATCACTCTTATCCAGACAGAGATTTGTGAATCTTTCATCTTCAACCAACGAGCTGTGG 1260
Db 400 PheGlnGln-----SerGlnThrCluThrValHisleuPheIleProAlaLeuSerValGly 418
QY 1261 GCCATCATCGGAGAGAGAGGAGGAGCAGATCAACAGCTGCGAGATTCGCGGAGCTCT 1320
Db 419 AlaIleIleGlyLysGlnGlyGlnHisIleLysGlnleuSerArgPheAlaGlyLaser 438
QY 1321 ATCAAGATTTGCCCTCGGAGAGGCGCCAGACGTCAGCGAAAGATGTGTATCATCAACG 1380
Db 439 lleLysIleAlaProAlaGlnAlaProAspAlaLysValArgMetValIleIleThrGly 458
QY 1381 CCACCGAAGCCGAGTTCAAGGCCAGGAGAGGATCTTTGGGAAATGAAAGAGAGAAAC 1440
Db 459 ProProGlnAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLysGlnLysasn 478
QY 1441 TTCTTTAACCCTCAAGAGAGAGTGAAGCTGAGAGCGCATATCAGAGTCCCTCTTCCACA 1500
Db 479 PheValSerProLysGlnGlnValLysleuGlnAlaHisIleArgValProSerPheAla 498
QY 1501 GCTGGCGGGGTGATTTGGCAAGAGTGGCAAGACCGTGAACGATTCGCAACTTAAACAGT 1560
Db 499 AlaGlyArgValIleGlyLysGlyGlyLysThrValasnGlnleuGlnasnleuSerSer 518
QY 1561 GCAGAAGTCAATCGGCTCTGACCAAAACCCGATGAAATGAGAAAGTATGTCGAGA 1620
Db 519 AlaGlnValValAlaProAlaGAspGlnThrProAspGlnasnspGlnValValLys 538
QY 1621 ATTATCGGCACTTCTTGTAGCCAGACTGCAAGCGCAAGATCAGGAAATGTGACAA 1680
Db 539 lleThrGlyHisPheTyrAlaCysGlnValAlaGlnArgLysIleGlnGlnIleleuThr 558
QY 1681 CAGGTAGAGCAGAGAGAGCAAA 1704
Db 559 GlnValLysGlnHisGlnGlnGln 566

RESULT 5
US-09-643-597-176
Sequence 176, Application US/09643597
Patent No. 6426072
GENERAL INFORMATION:
APPLICANT: Wang, Tonglong
APPLICANT: Fan, Liqun
APPLICANT: Kalos, Michael D.
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Hosken, Nancy
APPLICANT: Fanger, Gary R.
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Ajun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Henderson, Robert A.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

```

; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455c11
; CURRENT APPLICATION NUMBER: US/09/643,597
; CURRENT FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 369
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 176
; LENGTH: 579
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-643-597-176

Alignment Scores:
Pred. No.: 7,1e-157 Length: 579
Score: 1868.00 Matches: 368
Percent Similarity: 78.52% Conservative: 78
Best Local Similarity: 64.79% Mismatches: 96
Query Match: 31.74% Indels: 26
DB: Gaps: 6

US-09-270-437d-8 (1-3283) x US-09-643-597-176 (1-579)

QY 73 ATGAAACAGCTTTACATCGGAGCACTGAGCCCGCGTACCCGCGAGCACTCCGCGAG 132
Db 1 MetasnllyslentylrllleglyasnleusergluasnlaialaProserdyspleuugluser 20
QY 133 CTCTTTGGGAGACAGAGAGCTGCCCTGGGCGGACAGCTCTGTAAGTCCGGCTAACGC 192
Db 21 llephelyaspralalyserlleprovalserglyPropheluevallyserthrglytrala 40
QY 193 TTGCTGAGTACCCCGACAGAACTGGGCGCATCCGGCGCATGAGACCCCTCCGGGTAA 252
Db 41 phevalaspcybrospriuserltrialleuylalaileuylaleuuserglylys 60
QY 253 GTCGATTCGATGGGAAATGCAGAACTGATTCAGTCTCTAAAAGCTAAGAGC 312
Db 61 lleghleuhslelylsproillegluvalgluhiservalProlysarglnargile 80
QY 313 AGGAAATTCAGATTCGAAACATCCCTCTCAGCTGAGTGGAGGTGTGATGACATT 372
Db 81 Arglylsleuqlnleatrgasnlleprokonhleuqlntrpqlvalleuaserleu 100
QY 373 TTGGCTCATATGGGACAGTGGAGATGTGGAACAGTCAACAGACAGAAACCGCC 432
Db 101 leuvalglntylxglyalvalgluserCysgluqlnvalasnthrsperglutnra 120
QY 433 GTTGCAAGCTCAATATGCACAGAGAGAGAAATAAGCCATGAGAACTAAGC 492
Db 121 ValValasnvalthrlytserlserlyasprlnalarglnalaleuaprylsleuasn 140
QY 493 GGGCATCATTTGAGAACTACTCTTGAAGATTCTTCAATCCCGGATGAAAGGTGAGC 552
Db 141 GlpPheqlnleuqlnleuasnphethrleuylvalalalytrleprodarprglumetalala 160
QY 553 TCCCTCTCGCCCTCG-----CGAGCCAGCGT-----GGGACCACTCTTCCGG 600
Db 161 GlnqlnasnProleuqlnqlnProargglyargarglyleuqlnqlnarglylseser 180
QY 601 GAGCAAGCCAGCCCGCTGGGCGCACTTCAGGCGCAGACAGATGATTCGCGTCCG 660
Db 181 Argqlnqlnly--serProqlnlyserValserlysglnlyspocysaspleuProleuarg 199
QY 661 ATCCGTGCTCCACCCAGATTGTGTGGTCCATCATCGGAAAGAGGGCTTGAACCTAAG 720
Db 200 leuvalvalProthrqlnphelvalglyalallelleqlnlysglnqlnlyalathrlearg 219
QY 721 AACATCACTAAGCAGACCACTCCCGGTAGATATCAATAAAGAGAGACTGAGAGCT 780
Db 220 Asnlllethrlysglnthrglnsetrlyleasprvalhsarglylsghlnaspralglyala 239
QY 781 GCAGAGAGAGCTGTACCATCATGACCAACCCAGAGGGAGCTTGAAGCAATCCGCGATG 840
Db 240 Alaqlnlyserllethrilleuuserthrproqlnqlnlytnseralalacyslyser 259

```

```

QY 841 ATTCTGAATTCATGCAAGAGAGCAGATGAGACCAACTAGCCGAGAGATTCCTCTG 900
Db 260 lleuqlnilleuethlslysglnalglhsprileysphethrthrgluilleProleu 279
QY 901 AAATCTGGACACATATGCTGTGGTGAAGATGATTTGGAAGAAGAGCAAAATTTG 960
Db 280 lysilleuqlnleuasnphelvalglyargleuilleqlnlysglnlyarglyasnleu 299
QY 961 AAGAAATTTGAACATGAAACAGGACCAAGATTAACATTCATCTTTGACAGATTGAGC 1020
Db 300 lyslyslleqlnqlnserthrlyserthrlyserlethrleuserProleuqlnleuthr 319
QY 1021 ATATACACCCGGAAGAACCATCATCTGTGAAGGCGACAGTTGAGCGCTGCGCAGTCT 1080
Db 320 leuThrasnProqlnargthrillethrvallysglnvalglutnrcysalalysala 339
QY 1081 GAGATGAGATTATGAAAGAGCTGGTGAGGCTTTGAAATGATATGAGTGGCTGTAAAC 1140
Db 340 qlnqlnqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqln 359
QY 1141 ACCCACTCC-----GGATCTTC----- 1158
Db 360 leuqlnallehsleuilleProqlnleuasnleuasnalleuqlnleuPherProthr 379
QY 1159 -----TCCAGCTGTACCCCATACCACTTGGCCCGTTCGCGCAT 1200
Db 380 serqlnmeProProthrserglyProProseralameThrProProthrProqln 399
QY 1201 CATCACTCTTATCCAGAGAGAGAGATTTGATTCCTTCATCCCAACCGAGCTGTGAGC 1260
Db 400 pheqlnqln--serqlnthrqlnthrvalhsglnphelleproalaleuuseralgly 418
QY 1261 GCCATCATCGGAGAGAGGCGGACACATCAACAGCTGGCAGATTGCGCGAGCTCT 1320
Db 419 AlallelleqlnlysglnqlnqlnhslelysglnleuserArgphealaglylaser 438
QY 1321 ATCAAGATTGCCCTCGGAGAGCCCAAGCTCAGCGAAAGATGGTCATCATACCGG 1380
Db 439 llelyselealaproalaglnlaprodaralalyvalargmetvallellethrql 458
QY 1381 CCACCGGAGCCAGTTCAAGGCCCGAGGACGATCTTGGGAAACTGAAAGAGAAAC 1440
Db 459 ProProqlnlaqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqln 478
QY 1441 TTCTTAACCCCAAGAGATGAGCTGGAAGCCCATATCGAGTCCCTCTTCCACA 1500
Db 479 phevalserProlysglnqlnvallylsleuqlnallehslellyargvalProserpheala 498
QY 1501 GCTGGCCGGGTGATTGGCAAAAGGTGGCAAGCCGTGAACGAACCTGCAAGACTTAACAGT 1560
Db 499 AlaglyargvalilleqlnlysglylylysthrValasnqlnleuqlnleuqlnleuqln 518
QY 1561 GCAGAACTCATCTGCTGCTGCTGCAAAAGCCAGATGAAATGAGAAATGATGTCAGA 1620
Db 519 AlaglnvalvalProargasprqlnthrProasprqlnasnphqlnvalvallylys 538
QY 1621 ATTATCGGCACTTCTTGTGACCACTGCAACAGCGCAAGATCAGGAAATGTACAA 1680
Db 539 llethrqlnhsrlethrlyalacysqlnvalalaglnharglyslleqlnqlnleuqlnthr 558
QY 1681 CAGGTGAAGCAGAGAGAGAA 1704
Db 559 Glnvallysglnhsglnqlnqln 566

RESULT 6
US-09-480-884A-176
; Sequence 176, Application US/09480884A
; Patent No. 6482597
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongfong
; APPLICANT: Fan, Liqun
; APPLICANT: Hosken, Nancy A.

```

; APPLICANT: Kalos, Michael D.
 ; APPLICANT: Fanger, Gary R.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
 ; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
 ; FILE REFERENCE: 210121.45506
 ; CURRENT APPLICATION NUMBER: US/09/480,884A
 ; CURRENT FILING DATE: 2001-08-27
 ; NUMBER OF SEQ ID NOS: 330
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO: 176
 ; LENGTH: 579
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-480-884A-176

Alignment Scores:
 Pred. No.: 7,1e-157 Length: 579
 Score: 1868.00 Matches: 368
 Percent Similarity: 78.52% Conservative: 78
 Best Local Similarity: 64.79% Mismatches: 96
 Query Match: 31.74% Indels: 26
 Gaps: 6

US-09-270-437D-8 (1-3283) x US-09-480-884A-176 (1-579)

QY 73 ATGAACAAGCTTTATCATGAGGAGACCTGAGCCCGCCCTCACCGCCGACGACCTCCGCGAG 132
 Db 1 MetasulysleutyrlleglyAsnleuserglunsmAlaIaProserAspleugluser 20
 QY 133 CTCTTTGGGGGACAGAAAGCTCCCTGGCGGAGACAGCTCTGTGAATCGCGGTACGCC 192
 Db 21 lleheysapAlaIyAlleProValserGlyProPheleuVallysthrGlytAla 40
 QY 193 TTCTGAGTACCCCGGACGAGTGGGCGCATCGCGCCATCGGACGAGACCTTCGGGTAA 252
 Db 41 PheAlaspyspProsbpsgluserPrAlaleuysAlalleglunAlaleuSerGlylys 60
 QY 253 GTGGAATTGACATGGGAAATATCATGAAATTGATTACTACAGTCTTAAAGGCTAAAGAGC 312
 Db 61 lleglunehisgllyspProlleglunValglunhisserValProlysbarglunArglyle 80
 QY 313 AGGAAATTCAGATTGCAAAACATCTCTCTACCTGACAGTGGAGGAGGTGGATGAGCTT 372
 Db 81 ArglylsuenglInleargAsnleleProPheleuIntrpgluValleuAspserleu 100
 QY 373 TTGGCTCAATATGAGGACAGTGGAGATGTGGAACAATCAACACAGACAGAAACGCC 432
 Db 101 leuValglIntrgllyValValgluserCysglunValAlasnthraspsersgluthrAla 120
 QY 433 GTTGTCAAGCTCATATGACACAGAGAGAAAGCAAAATWAGCCATGAGAGAGCTTACG 492
 Db 121 ValValAsnValthrlyrserSerlysbpsglInAlarglInAlaleuAspLySleuAsn 140
 QY 493 GGGCATCATGTTTGAAGACTACTCTTCAGATTCTTCACTCCCGATGAGAGAGTACG 552
 Db 141 GlyPheglInleuInsbPheThrleuysValAlaTyrleleProAspglunMetAlaIa 160
 QY 553 TCCCTTGGCCCTCCAG-----CGAGCCACAGCT-----CGGACCACTTCTCCCG 600
 Db 161 GlInghAsnProleugInglInProArgGlyArgArgGlyleuGlyglInArgGlyser 180
 QY 601 GAGCAAGGCCACCGCCCTGGGGGCACTTCTCAGCGGACAGACAGATTGATTTCCGCGCG 660
 Db 181 ArgGInglY---serProglYserValserlysgInlyserProCysAspleuProleuAsn 199
 QY 661 ATCTGTGTCCCAACCCAGTTTGTGTGTCATCATCGGAAAGAGAGGCTTGAACATTAAG 720
 Db 200 leuValProthrGlnPheValGlyAlaIlelleGlylysgInglYAlaThrleArg 219
 QY 721 AACATCATAGAGAGACCGAGTCCCGGATGATATCATAGAAAGAAAGAACTTGGAGCT 780
 Db 220 AsnleThrlysgInthrGlnserlyslleAspValHisArglysgInunsmAlaglyAla 239

QY 781 GCAGAGAGCCCTGTGCACCATTCATGCCACCCGAGGAGGAGCTTGTGAGCATGCCGATG 840
 Db 240 AlaGlyuserleThrleleuserThrProglunGlyThrSerAlaIaCyslyser 259
 QY 841 ATCTTGAATCATCAGAAAGAGGACAGTGAACCAACTGACCGGAGAGATTCCTCG 900
 Db 260 lleleuGlnleleuHislelysgInAlaInsbPheleuIlelysgInlyArgAsnleu 279
 QY 901 AAAATCTTGGCACACAAATGGCTTGTGGAAGATGATTTGAAAAGAGGAGAAATTTG 960
 Db 280 lyslleleuAlaHisAsnAsnPheValGlythrleuIleGlylysgInlyArgAsnleu 299
 QY 961 AAGAAATTTGAACATGAAACAGGAGACCAAGATTAATCATCTTGTGAGGATTTGAC 1020
 Db 300 lyslyslleGlnInAspThrAspThrlyslleThrlyleleProleugInleuThr 319
 QY 1021 ATATCAACCCCGGAAAGAACCATCATCTGTGAAGGCGACAGTTGAGCCTTGCCAGTGT 1080
 Db 320 leuYrAsnProglunbArgThrleThrVallysglyAsnValgluthrCysAlaIysAla 339
 QY 1081 GAGATGAGATTATGAAAGACCTGCGTGAAGCCTTGTGAATATGATGCTGCTTTAC 1140
 Db 340 GlunGlnGlnleuMetlysllyleArggluserThrGlnAsnAspIleAlSerMetAsn 359
 QY 1141 ACCCACTCC-----GGATACCTG----- 1158
 Db 360 leuGlnAlaHisleuIleProglYleuAsnleuAsnAlaleuGlyleuPheProthr 379
 QY 1159 -----TCCAGCCTGACCCCATACCAAGTTTGCGCCCTTCCCGAT 1200
 Db 380 SerGlyMetProProthrInSerGlyProProserAlaMetThrProProtyrProglIn 399
 QY 1201 CATCATCTTATCCAGACAGAGATTTGATCTTTCATCCCAACCCAGAGCTGTGGCG 1260
 Db 400 PheglunIn---serGluThrIngluthrValHisleInhelleProAlaleuSerValGly 418
 QY 1261 GCCATCATCGGAGAAAGGCGGCGACACATCAACACAGCTGGCGAGATTGCGCGAGCTCT 1320
 Db 419 AlallelleGlylysgInglYglnHislelelysgInleuserArgPheAlaGlyAlaser 438
 QY 1321 ATCAAGATTGCCCTCGGAGAGGCCGAGAGTCAAGCAAGATGTCATCATATCAACGGG 1380
 Db 439 llelyslleAlaProAlaGlnAlaProAspAlaIysValArgMetValleleThrGly 458
 QY 1381 CCAACCGGAGCCAGTTCAAGGCCACAGGACGATCTTGTGGAAATCGAAAGAGAAAC 1440
 Db 459 ProProglunAlaGlnPheysAlaGlnGlyArgGlyleThrGlylysllelysgInunsn 478
 QY 1441 TTCTTAACCCCAAGAGAGAGTGAAGCTGGAAGCCGATATCAAGATGCCCTTTCACA 1500
 Db 479 PheValSerProlysgInglunVallySleuGlnAlaHisleleArgValProserPheAla 498
 QY 1501 GCTGCGCGGGTGAATTGGCAAGGTGGCAAGACCGTGAACGACTGCAGAACTTAAACAGT 1560
 Db 499 AlaGlyArgValalleGlylysglyGlylyslThrValAsnleuInleuGlnAsnleuser 518
 QY 1561 GCAGAGTCACTGCTGCTGTGAACCAAGCCAGATGAAAAATGAGAAAGATGCTGTGAGA 1620
 Db 519 AlaGlnValValValProArgAspGlnThrProAspGlnAsnAspGlnValVallyS 538
 QY 1621 ATTATGGGCACTTCTTGTGACCAAGACTGCACAGCGCAAGATCAAGGAATGTGAACA 1680
 Db 539 lleThrGlyHisPheThrAlaCysGlnValAlaGlnArglyslleGlnGlnleleuThr 558
 QY 1681 CAGGTGAGCAGAGAGCAAGAA 1704
 Db 559 GlnVallysgInHisGlnGlnIn 566

RESULT 7
 US-09-542-615A-176
 ; Sequence 176, Application US/09542615A
 ; Patent No. 6518256
 ; GENERAL INFORMATION:

APPLICANT: Wang, Tonglong
 APPLICANT: Fan, Liqun
 APPLICANT: Kalos, Michael D.
 APPLICANT: Bangur, Chaelanya S.
 APPLICANT: Hosken, Nancy A.
 APPLICANT: Fanger, Gary R.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
 TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
 FILE REFERENCE: 210121.455C8
 CURRENT APPLICATION NUMBER: US/09/542.615A
 CURRENT FILING DATE: 2000-04-14
 NUMBER OF SEQ ID NOS: 350
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 176
 LENGTH: 579
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-542-615A-176

Alignment Scores:
 Pred. No.: 7,1e-157 Length: 579
 Score: 1868.00 Matches: 368
 Percent Similarity: 78.52% Conservative: 78
 Best Local Similarity: 64.79% Mismatches: 96
 Query Match: 31.74% Indels: 26
 DB: 4 Gaps: 6

US-09-270-437d-8 (1-3283) x US-09-542-615A-176 (1-579)

QY 73 ATGACCAAGCTTTATCATGGGAGACCTGAGCCCGCGTACCGCGAGCACTCGGCGAG 132
 Db 1 Metamylsleutrylleghyasnleusergluasnlalalproserhspleuuser 20
 QY 133 CTTCTGGGAGAGAGAGCTGCCCTGGCGGAGAGCTGCTGTAAGTCCGCTTACGCC 192
 Db 21 llephelyasplalalyslleprovalserglypropheluevallythrlytrala 40
 QY 193 TTGCTGACTACCCGACCAAGAACTGGGCTATCCGCGCATCGAGACCCCTCGGGTAA 252
 Db 41 PhevalaspCysProaspgluserltpalaleuylsalaleglualaleuserglyls 60
 QY 253 GTCGATTCGATGGGAAATCATGAAAGTATTACTAGCTCTCTAAAAAGCTAAGAGC 312
 Db 61 lleghleuhslelylsproillegluvalgluhlservalprolysharglnaglie 80
 QY 313 AGGAAATTCAGATTGGAACATCCCTCTCACCTGACGTGAGGAGGTGTGATGACTT 372
 Db 81 Arglylsleuqlnleltharganilleproprohlsleuqlntrpgluvalleuaspserleu 100
 QY 373 TTGGCTCAATATGAGACAGTGAAGATGTGGACCAAGTCAACACAGACACAGAAACCGCC 432
 Db 101 leuvalglntlyrlyalvalgluserCysgluglnvalasnthraspsergluthala 120
 QY 433 GTTGTCAAGCTCATATGACACAGAGAGAAATAAGCATGAGAGAACTAAGC 492
 Db 121 Valvalasnvalthrlytserseylsaspglnalarglnalaleuaplylsleuasn 140
 QY 493 GGCATCATGTTGGAACCTACTCTTCAAGATTTCTCAATCCCGATGAAGAGGTAGC 552
 Db 141 GlpHehlnleuqlnleuasnphethrleuylvalalalyrilleproaspglumetalala 160
 QY 553 TCCCTTCGCCCCCTCAG-----CGAGCCAGCGT-----GGGAGACACTTCCCGG 600
 Db 161 GlnglnasnproleuglnglnlnproarglylArgrarglyleuqlnglnarglylsSer 180
 QY 601 GAGAGAGCCACGCGCCCTGGGAGCACTTCAAGCCAGACAGATGATTTCCCGCTGCGG 660
 Db 181 Argnglyl--SerProglyServalserlysglnlysproCysapleuProleuarg 199
 QY 661 ATCTGTGTCACCAAGATTGTGTGTCATCATCGAAAGAGGCTTGAACATAAG 720
 Db 200 leuvalvalprothrlnphevalglYalalrlelleghlysglnlylAlatThrlearg 219

QY 721 AACATCAATAGAGACCCAGTCCGGGTAGATTCATAGAAAAGAACTCGAGCT 780
 Db 220 AsnlethrlysglnthrlnserlyleaspvalhsarglylsGluasnlalglYala 239
 QY 781 GCAGAGAGCTGTCACCATTCACAGCCACAGAGGCACTTCTGAAGATGCGCATG 840
 Db 240 Alaglylsseerillethrleuaserlthproaluglythrseralalaclylseser 259
 QY 841 ATTCTGAATCATGACGAAAGAGAGATGACCAAACTAGCCGAAAGATTCCTCTG 900
 Db 260 lleuqlnleuethslslysglnalgluaspllelysphehrnglnulleproleu 279
 QY 901 AAAATCTTGCAACAATGCTGTTGTAAGATCATGATGGAAGAAAGAGCAAAATTG 960
 Db 280 LyslleuvalahlsasnasnphvalglYargyleuileghlylsGlnlylYarglsneu 299
 QY 961 AAGAAATTTGAACATGAAACAGGAGCAAGATTAACATCTCATCTTGACAGATTGAGC 1020
 Db 300 LysylsilleghlylnaspthraspthrlysllethrlyleserProleuglnleuthr 319
 QY 1021 ATATCAACCCGAAAGAACCATCATCTGTGAAGGCGACAGTTGAGGCTGTGCCAGTCT 1080
 Db 320 leutyrasnprogluargThrillethrvallysglyasnvalgluthrCysAlalysla 339
 QY 1081 GAGATGAGATTATGAAGAGCTGCGTGAAGCCTTTGAAAATGATATGCTGCTGTAC 1140
 Db 340 Gluglngluilemetylelylslearglnserlyrgluasnspllelasermetasn 359
 QY 1141 ACCCACTCC-----GGATCTTC----- 1158
 Db 360 leuqlnlahlsleuilleproglyleuasnleuasnlaaleuqlyleuhsleprothr 379
 QY 1159 -----TCCAGCTGTACCCCATCACCAAGTTGGCCCGCTTCCGCGAT 1200
 Db 380 SerglymeuProProthrserglyProProseralameThrProProtyrProgin 399
 QY 1201 CATCACTTTATCCAGAGAGAGATGTGATCTCTTATCCCAACCAAGGCTGGGCG 1260
 Db 400 Pheglugln--Serlythrlythrvalhslnphelleprohaleuservalgly 418
 QY 1261 GCCATCATCGGAGAGAGGCGCACATCAAAACAGCTGCGAGATTCGCGGAGCTCT 1320
 Db 419 Alalrlelleghlylsnglnlylnhlslelylsnglnleuserlthpheelaglylaser 438
 QY 1321 ATCAAGATTGCCCTCGGAGGCGCCAGACGTGACGAAAGATGTGATCATCACCGGG 1380
 Db 439 llelysllealaprohlaqlunlaProaspalalyvalargmetvalleillethrgly 458
 QY 1381 CCACCGGAGCCCATGTTCAAGGCCCGAGGACCGATCTTTGGGAAACGAAAGAGAAAC 1440
 Db 459 ProProgluhalaglnpheylsalainglnlylYarglylelylslysllelysglnluasn 478
 QY 1441 TTCTTAAACCCCAAGAGAGAGTGAAGCTGGAAGCGCATCATCAAGTGCCTCTTACA 1500
 Db 479 PhevalserProlysglnlylvallylsleuqlnlahlslelthargvalProserPheala 498
 QY 1501 GCTGCGCGGCTGATTGCAAAAGGTGCGAAGACCGTGAACGAATCTGCAACTTACAGT 1560
 Db 499 AlaglylYargvalilleghlylsGlnlylsThrValasnglueuqlnleuasnleuser 518
 QY 1561 GCGAAGATCATGCTGCTCGTGAACCAAGCCAGATGAAATGAGAAAGTATGATCTGACA 1620
 Db 519 AlaglylvalalvalProargaspglntthProaspgluasnaspglvalvalVallys 538
 QY 1621 ATTATGGGGACCTTTGTGACGACATGACAGGCGCAAGATCAGGAAATTTGACAA 1680
 Db 539 llethrlylnhspeytralaCysglnvalAlaglnlrgylsilleghnglnulleuthr 558
 QY 1681 CAGGTGAAGCAGACAGACAGAA 1704
 Db 559 Glnvallysglnhsglnglngln 566

RESULT 8

```
US-09-606-421B-176
; Sequence 176, Application US/09606421B
; Patent No. 653315
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.45509
; CURRENT APPLICATION NUMBER: US/09/606,421B
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 176
; LENGTH: 579
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-606-421B-176

Alignment Scores:
Pred. No.: 7.1e-157 Length: 579
Score: 1868.00 Matches: 368
Percent Similarity: 78.52% Conservative: 78
Best Local Similarity: 64.79% Mismatches: 96
Query Match: 31.74% Indels: 26
Gaps: 6

US-09-270-437D-8 (1-3283) x US-09-606-421B-176 (1-579)

QY 73 ATGAACAAGCTTACATGGGAACTGAGCCCGCCGTCACCGCCGAGACCTCCGCGAG 132
Db |||||
QY 1 MetasnysLeuYrIleGIyAsnLeuSerGIuAsnAlaIaProSerAspLeuGIuSer 20
Db |||||
QY 133 CTCCTTGGGGACAGAAAGCTGCGCCCTGCGGGACAGAGCTCTGCTGAAGTCCGGCTACGCC 192
Db ::::|
QY 21 IleheYsAspAlaYsIleProValSerGIyProheLeuValYsThrGIyTYrAla 40
Db |||||
QY 193 TTCGTGAGTACCCCGACCAAGAACTGGGCGCATCGCGCCATCGAGACCTCTGGGGTAA 252
Db |||||
QY 41 PheValAspCysProAspGIuSerTyrAlaLeuYsAlaIleGIuAlaLeuSerGIyLys 60
Db |||||
QY 253 GTGAATTGATGGGAAATATCAAGAACTGATTACTCACTCTCTAAAGCTAAGGAGC 312
Db ::::|
QY 61 IleGIuLeuHISGIyLysProIleGIuValGIuHISerValProLysArgGIuArgIle 80
Db |||||
QY 313 AGGAATATTCAGATTGGAACATCCCTCCCTCACCTGAGAGGGAGGTGGATGAGCTT 372
Db |||||
QY 81 ArgLysLeuGIuIleArgAsnIleProIleHisLeuIleThrGIuValLeuAspSerLeu 100
Db |||||
QY 373 TTGGCTCAATATGAGAGAGTGGAGAACTGTGAAACAAGTCAACAAGACAGAAACCGGC 432
Db |||||
QY 101 LeuValGIuIleTyrGIyValValGIuSerCysGIuGIuValAsnThrAspSerGIuThrAla 120
Db |||||
QY 433 GTTGTAAAGCTCAATATGCAACAAGAAAGAAATAGCAATGACAGAGTAAAGC 492
Db |||||
QY 121 ValValAsnValThrTyrSerSerLysAspGIuAlaArgGIuAlaLeuAspLysLeuAsn 140
Db |||||
QY 493 GGGCATCAGTTGAGAACTCTCTTCAGAACTTCTTACATCCCGGATGAGAGGTGAGC 552
Db |||||
QY 141 GlyPheGIuIleuGIuAsnPheThrLeuLysValAlaTyrIleProAspGIuMetAlaAla 160
Db |||||
QY 553 TCCCTTCGCCCGCCCTGAG-----CGAGCCCAAGCGT-----GGGACCACTCTTCCGG 600
Db ::::|
QY 161 GIuGIuAsnProLeuGIuGIuIleProArgGIyArgArgGIyLeuGIyGIuArgGIySer 180
Db |||||
QY 601 GAGCAAGGCGCAGCGCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGG 660
Db |||||
```

```
Db |||||
QY 181 ArgGIuGIy----SerProLysSerValSerLysGIuLysProCysAspLeuProLeuArg 199
Db |||||
QY 661 ATCTGGTCCCAACCCAGTTTGTGGTCCATCATCGAAAGAGAGGAGCTTGACCATTAAG 720
Db ::::|
QY 200 LeuLeuValProThrGIuIlePheValGIyAlaIleIleGIyLysGIuGIyAlaThrIleArg 219
Db |||||
QY 721 AACATCATTAAGACACCCAGTCCCGGGTAGATTCATAGAAAGAAAGAACTGGAGCT 780
Db |||||
QY 220 AsnIleThrLysGIuIleThrGIuIleSerLysIleAspValIleArgLysGIuAsnIleGIyAla 239
Db |||||
QY 781 GCAGAGAGCCGTGATCCATCCATGCCACCCCAAGGGGAGCTTCTGAACATGCGCATG 840
Db |||||
QY 240 AlaGIuLysSerIleThrIleLeuSerThrProGIuGIyThrSerAlaIleCysLysSer 259
Db |||||
QY 841 ATCTTGAATCATGCAGAAAGAGGAGAGTGAACCAAACTAGCCGGAAGAGTTCCTGCG 900
Db |||||
QY 260 IleLeuGIuIleMetHISLysGIuAlaGIuAspIleLysPheThrGIuGIuIleProLeu 279
Db |||||
QY 901 AAAATCTTGGCACACAAATGGCTTGGTGAAGACTGATTGGAAGAAAGAGCAAAATTGG 960
Db |||||
QY 280 LysIleLeuAlaHISAsnAsnPheValGIyArgLeuIleGIyLysGIuGIyArgAsnLeu 299
Db |||||
QY 961 AAGAAATATGAACATGAACAGAGGACCAAGATTAACATCTCTTTCAGAGATTGAGC 1020
Db |||||
QY 300 LysLysIleGIuGIuAspThrAspThrLysIleThrIleSerProLeuGIuGIuLeuThr 319
Db |||||
QY 1021 ATATTAACAACCCGGAAGAAACCATCATCTGTGAAGGGACAGATTGAGCGCCGAGTGT 1080
Db ::::|
QY 320 LeuTyrAsnProGIuArgThrIleThrValLysGIyAsnValGIuThrCysAlaLysAla 339
Db |||||
QY 1081 GAGATTAAGATTATGAAGAGCTCGTAGGCGCTTTGAAATGATATGATCGCTGTATAC 1140
Db |||||
QY 340 GIuGIuGIuIleMetLysLysIleArgGIuSerTyrGIuAsnAspIleAlaSerMetAsn 359
Db |||||
QY 1141 ACCCACTGC-----GCATCTTC----- 1158
Db |||||
QY 360 LeuGIuAlaHISLeuIleProGIyLeuAsnLeuAsnAlaLeuGIyLeuPheProProThr 379
Db |||||
QY 1159 -----TCAGGCTTGATCCCGCCCATCAACAGTTTGGCCGCTTCCCGCAT 1200
Db |||||
QY 380 SerGIyMetProProProThrSerGIyProProSerAlaMetThrProProTyrProGIu 399
Db |||||
QY 1201 CATCATCTTATATCCAGACAGAGAGATTGAAATCTTCAATCCCAACCCAGAGCTGTGGC 1260
Db |||||
QY 400 PheGIuGIuIle--SerGIuThrGIuThrValHISGIuPheIleProAlaLeuSerValGIy 418
Db |||||
QY 1261 GCCATCATCGGGAAAGAGGGGGACACATCAACAGCTGGCGAGATTGCCGGAGCTCT 1320
Db |||||
QY 419 AlaIleIleGIyLysGIuGIuIleHISLysGIuIleLysGIuLeuSerArgPheAlaGIyAlaSer 438
Db |||||
QY 1321 ATCAAGATTGCCCCGCGGAAGGCCAGAGCTGACGGAAAGATGGTCATCATCAACCGGG 1380
Db |||||
QY 439 IleLysIleAlaProAlaGIuAlaProAspAlaLysValArgMetValIleIleThrGIy 458
Db |||||
QY 1381 CCAACCGGAAGCCAGTTCAAGGCCAGAGCGAGATCTTGGGAAATCTGAAGAGAAAC 1440
Db |||||
QY 459 ProProGIuAlaIleGIuPheLysAlaGIuGIyArgIleTyrGIyLysIleLysGIuGIuAsn 478
Db |||||
QY 1441 TTCTTAACCCCAAGAAAGAGTGAAGCTGGAAGCGCATATCAAGAGTGCCTTCTTCACA 1500
Db |||||
QY 479 PheValSerProLysGIuGIuValLysLeuGIuAlaHISLysArgValProSerPheAla 498
Db |||||
QY 1501 GCTGGCCGGGTGATTTGGCAAGAGGTGCAAGACCGTGAAGAACTGCAAGAACTTAACAGT 1560
Db |||||
QY 499 AlaGIyArgValIleIleLysGIyGIyLysTyrValAsnGIuLeuIleAsnLeuSerSer 518
Db |||||
QY 1561 GCAAGAGTATGTGCTGTGTAACCAACCGCAAGATGAAGAAATGAGAAATGATGCTCAG 1620
Db ::::|
QY 519 AlaGIuValValValProArgAspGIuThrProAspGIuAsnAspGIuValValValLys 538
Db |||||
QY 1621 ATTATCGGCACTTCTTGTGCTGAGCAAGACTGACAGCGGAAGATCAGGGAATTTGACAA 1680
Db |||||
```


[illegible]

Db 494 -----ProGlyProAlaProIseGlyProProAlaProGlyProAlaProGlnGly 509
QY 1920 GGGGGGAAGTTCAGCCAGGTTTGCAGAA-----CCACCGAGCCCCGCC 1964
Db 510 TrpGlyAsnAlaGlyProHisTrpGlnGlnGlnAlaProProAspProAla 526

RESULT 11
PCT-US94-01782-2
Sequence 2, Application PC/TUS9401782
GENERAL INFORMATION:
APPLICANT: THE GOVERNMENT OF THE UNITED STATES
APPLICANT: AS REPRESENTED BY THE SECRETARY, DEPARTMENT OF
APPLICANT: HEALTH AND HUMAN SERVICES
TITLE OF INVENTION: NOVEL FUSE BINDING
TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEMAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01782
FILING DATE: 22-FEB-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/021,608
FILING DATE: 22-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAM S. FEILER
REGISTRATION NUMBER: 26,728
REFERENCE/DOCKET NUMBER: 2026-4063PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 644
TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: Unknown
MOLECULE TYPE: Peptide/Protein
HYPOTHETICAL: No
ORIGINAL SOURCE:
ORGANISM: Human
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE: HL60
ORGANELLE:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
OTHER INFORMATION: Amino Acid 149 (Xaa) is Met or Ile
PCT-US94-01782-2

Alignment Scores:
Pred. No.: 1,52e-14 Length: 644
Score: 263.50 Matches: 140
Percent Similarity: 37.73% Conservative: 86

Best Local Similarity: 23.37% Mismatches: 237
 Query Match: 4.48% Indels: 136
 DB: 5 Gaps: 25

US-09-270-437D-8 (1-3283) x PCT-US94-01782-2 (1-644)

```

QY 336 CCTCTCTCAGCTTCACTGGAGAGGTGTGTGATGAGCTTTGGCTCAATATATGGACAGTGA 395
   |||||
Db 10 ProSerSerGlySerIaGlyGlyGly-----GlyGlyGly 22
QY 396 GAATGTGAACAAGTACACACACACACACACACACACACACACACACACACACACACAC 455
   |||||
Db 23 GlyGlyGly-GlyValAsnAspAlaPheYsAspAlaLeuGlnArg-----AlaAr 39
QY 456 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 515
   |||||
Db 39 gGlnIleAlaIleAlaIleGlyGlyAspAlaGlyThrSerLeuAsnSerAsnAspIleI 59
QY 516 CTTCAGAGATTTCTTCATCCGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 575
   |||||
Db 59 YTyGlyGlyGlnYsArgProLeuGlnAspGlyAspGlnProAspAlaLeuValAla 79
QY 576 CAGAGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 635
   |||||
Db 79 AProGlnAsnSpSerPheGlyThrGlnLeu-----ProPheMetHisGlnGlnIle 97
QY 636 CAGACAGATGATTTCCGCTGCGAGATCCGTGGTCCCAAGAGAGAGAGAGAGAGAGAGAG 695
   |||||
Db 97 rArgSerVal---MetThrGlnGlnIleValAlaProAspGlyMetValGlyPheIleI 116
QY 696 CGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 755
   |||||
Db 116 eGlyYrGlyGlyGlnGlnIleSerArgIleGlnGlnIleSerGlyGlyIleGlnIle 136
QY 756 CCATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 815
   |||||
Db 136 e---AlaProAspSerGlyGlyLeuProGlnArgSerCys**LeuThrGlyThrProG 155
QY 816 GGGAGCTTCTGACATGCGCATGATCTTGAATCATGACAGAGAGAGAGAGAGAGAGAGAG 861
   |||||
Db 155 useValGlnSerIleAlaYsArgLeuLeuAspGlnIleValGlnYsGlyArgProAlaPr 175
QY 862 -----GAGCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 911
   |||||
Db 175 oGlyPheHisHisGlyAspGlyProGlyAsnAlaValGln-----GluIleMetI 192
QY 912 ACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 971
   |||||
Db 192 eProAlaSerIleYsAlaGlyLeuValIleGlyGlyGlyGlyGlnThrIleYsGlnLeuG 212
QY 972 ACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1031
   |||||
Db 212 nGlnArgAlaGlyAlaIleYsMetValMet-----IleGlnAsp-----GlyPr 226
QY 1032 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1076
   |||||
Db 226 oGlnAsnThrGlyAlaAspIlyProLeuArgIleThrGlyAspProIlyrIlyValGlnG 246
QY 1077 TGCTGAGATGAGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1136
   |||||
Db 246 nAlaYsGlnMetValLeuGlnLeuIleArgAsp----- 257
QY 1137 TAAACCCACTCCGAGTACTTCTCCAGCTGTACCCCATCACACAGTGTGGCCCGTCC 1196
   |||||
Db 258 -----GlnGlyGlyPheArgGlnVal-----ArgAsnGlnIlyrIly----- 269
QY 1197 GCATCACTCACTTATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1256
   |||||
Db 270 -----SerArgIleGlyYsGlnGlnIleAspValProIleProArgPheAlaVala 287
QY 1257 GGGGCGCATATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1316
   |||||
Db 287 lGlyIleValIleGlyYrArgAsnGlyGlnMetIleYsIleGlnAsnAspAlaGlyVa 307
  
```

```

QY 1317 CTCTATCAAGATTGCCCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1376
   |||||
Db 307 lArgIleGlnPheYsProAspAspIly---ThrThrProGlnArgIleAlaGlnIleTh 326
QY 1377 CGGGCCAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1407
   |||||
Db 326 rGlyProProAspArgCysGlnHisAlaIleGlnIleIleThrAspLeuArgSerVa 346
QY 1408 ----- 1425
Db 346 lGlnAlaGlyAsnProGlyGlyProGlyPheGlyArgGlyYrGlyValGlnG 366
QY 1426 -----CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1457
   |||||
Db 366 YAsnTrpAsnMetGlyProProGlyGlyLeuGlnGlnPheAsnPheIle----- 382
QY 1458 AGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1517
   |||||
Db 383 -----ValProThrGlyYsThrGlyLeuIleIleG 393
QY 1518 CAAAGGTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1577
   |||||
Db 393 YLysGlyGlyGlnThrIleYsSerIleSerGlnGlnSerGlyAlaArgIleGlnLeuG 413
QY 1578 TCGTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1632
   |||||
Db 413 nArgAsnProProProAsnAlaAspProAsnMetIlySerPheThrIleArgIlyThrPr 433
QY 1633 -----TTCTTGTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1682
   |||||
Db 433 oGlnGlnIleAspYrAlaAspGlnLeuIleGlnIleYsIleGlyGlyProValAsnPr 453
QY 1683 GGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1742
   |||||
Db 453 oLeuGlyProProValProHisGlyProHisGlyVal-ProGlyProHisGlyProProG 473
QY 1743 CTCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1802
   |||||
Db 473 lYProProGly-ProGly-----ThrPro-----Met 481
QY 1803 AGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1862
   |||||
Db 482 GlyProIlyrAsnProAlaProIlyrAsnProGlyPro----- 493
QY 1863 CTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1919
   |||||
Db 494 -----ProGlyProAlaProHisGlyProProAlaProIlyrAlaProGlnGly 509
QY 1920 GGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1964
   |||||
Db 510 TrpGlyAsnAlaIlyrProHisTrpGlnGlnAlaIleProProAspProAla 526
  
```

RESULT 12
 US-08-021-608D-10
 ? Sequence 10, Application US/08021608D
 ? Patent No. 5580760
 ? GENERAL INFORMATION:
 ? APPLICANT: LEVENS, DAVID L., DUNCAN,
 ? APPLICANT: ROBERT C., AND AVIGAN, MARK I.
 ? TITLE OF INVENTION: NOVEL FUSE BINDING
 ? NUMBER OF SEQUENCES: 24
 ? CORRESPONDENCE ADDRESSES:
 ? ADDRESSEE: MORGAN & FINNEGAN
 ? STREET: 345 PARK AVENUE
 ? CITY: NEW YORK
 ? STATE: NEW YORK
 ? COUNTRY: USA
 ? ZIP: 10154
 ? COMPUTER READABLE FORM:
 ? MEDIUM TYPE: FLOPPY DISK
 ? COMPUTER: IBM PC COMPATIBLE

QY 1743 CTCCACAGGACCCAGCAAAACAGGATGATGATGCCCTTCCACACCTGACAGAAATG 1802
|||
Db 472 lYProProgly-Progly-----ThrPro-----Met 480
QY 1803 AGACCAAGCGACGCCAGATCGGAGCAACCAAGACCATCTGAGGATGAGAAAT 1862
|||
Db 481 GlyProTyraaProAlaProTyraaProglyPro----- 492
QY 1863 CTGCGAGGCGGCGGACCTCTGCGAGG---CCCTGAGAACCCGAGGCGCGAGAGG 1919
|||
Db 493 -----ProglyProAlaProHisGlyProProAlaProTyraaProgly 508
QY 1920 GCGCGGGAAGTCTGACGCGAGTTTCCAGAA-----CCACCGAGCCCGCC 1964
|||
Db 509 TpgGlyAsnaIaTyProHisTpgInGlnIaIaProProAspProAla 525

RESULT 13
US-08-726-160-10
Sequence 10, Application US/08726160
Patent No. 5734016
GENERAL INFORMATION:
APPLICANT: LEVENS, DAVID L., DUNCAN,
APPLICANT: ROBERT C., AND AVIGAN, MARK I.
TITLE OF INVENTION: NOVEL FUSE BINDING
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/726,160
FILING DATE: 04-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/021,608
FILING DATE: 22-FEB-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAM S. FELTER
REGISTRATION NUMBER: 26,728
REFERENCE/DOCKET NUMBER: 2026-4063US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 751-6840
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 643
TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: Unknown
MOLECULE TYPE: Peptide/Protein
HYPOTHETICAL: No
ORIGINAL SOURCE: Human
CELL LINE: HL60
FEATURE:
OTHER INFORMATION:
OTHER INFORMATION: Amino Acid 148 (Xaa) is Met or Ile
US-08-726-160-10

Alignment Scores: 2,06e-14 Length: 643
Pred. No.: 262.00 Matches: 140
Score:

Percent Similarity: 37.56% Conservative: 85
Best Local Similarity: 23.37% Mismatches: 237
Query Match: 4.45% Indels: 137
DB: 1 Gaps: 25

US-09-270-437D-8 (1-3283) x US-08-726-160-10 (1-643)

QY 336 CCTCTCTCACTTCGAGTGGAGGTGTGGATGACTTTTGGCTCAATATATGACAGTGA 395
|||
Db 10 ProserSerGlySerAlaGlyGlyGly-----GlyGlyGly 22
QY 396 GAATGTGAGCAAGTCAACACAGACAGAACCCGCTTGTCACATGATGCAAC 455
|||
Db 23 GlyGlyGly-GlyValAsnaSpAlaPheIysAspaIaIeuGlnArg-----AlaAr 39
QY 456 AAGAGAAAGCAAAATATAGCATGAGAGAGTAAAGCGGAGCTAGTTGGAAGACTCTC 515
|||
Db 39 GlnIleAlaIaIaIysIleGlyGlyAspaIaGlyThrSerIeuAsnaSerAsnaSpIyl 59
QY 516 CTTCAGATTTCTTACATCCGGATGAAGAGGTGAGTCCCTTGCCCTTCAGCGAGC 575
|||
Db 59 YTyrcIyGlyGlnIySaIyGProIeuGlnAspGlyAspGlnProAspaIaIyIyValAl 79
QY 576 CCAAGCTGGGACCACTCTCCCGGAGCAAGGCCACCCCTTGCGGCACTTCTCAGC 635
|||
Db 79 aProGlnAsnaSpSerPheGlyThrGln-----LeuProProwethIsgInGlnI 96
QY 636 CAGACAGATTGATTTCCGCTGCGGATCCTGAGTCCGCCACCAAGTTTGTGGTCATAT 695
|||
Db 96 nArgSerVal---MetThrGlnIyGlyIyValaIyProAspGlyIyValaIyGlyPheIleI 115
QY 696 CGAAAGAGGCGCTTGAACCATTAAGAACATCACTAAGACAGACCGAGTCCGAGTATAT 755
|||
Db 115 eGlyArgGlyGlyGlnIleSerArgIleGlnGlnIleSerArgIySerIleGlnI 135
QY 756 CCATGAAAGAGAACTCTGAGAGTGCAGAGAACCTGTCAACATCCATGCGACCCAGCA 815
|||
Db 135 e---AlaProAspSerIyGlyIyLeuProGlnArgSerCys***LeuThrGlyThrProG 154
QY 816 GGGGACTTCTGAAGCAGTCCGATGATCTTGAATCATGACAGAA----- 861
|||
Db 154 uSerValGlnSerAlaIySaIyGleuAspGlnIleValaIyGlnIySaIyArgProAlaPr 174
QY 862 -----GAGGACATGAGACCAAACTAGCCGAGAAATTCCTGAAATCTTGGC 911
|||
Db 174 oGlyPheIleHisGlyAspGlyProGlyAsnaIaIaValGln-----GlnIleMetI 191
QY 912 ACACAAATGCTTGGTGGAGACTGATTTGAAAAGAAAGCAGAAATTTGAAGAAATGA 971
|||
Db 191 eProAlaSerIySaIyGlyIyLeuValIleGlyIySaIyGlyIyGlnIleIyGlnIleuG 211
QY 972 ACATGAACAGAGGACCAAGATTAACAAATCTATCTTTGAGAGATTTGAGCATATACACC 1031
|||
Db 211 nGlnIyArgAlaGlyIyValIyIyMetValMet-----IleGlnAsp-----GlyPr 225
QY 1032 GGAAGAACCC-----ATCACTGTAAGAGGACAGCTTGAAGCTGTGGCAG 1076
|||
Db 225 oGlnAsnThrGlyAlaAspIySerProIeuArgIleThrGlyAspProTyrlIyValaIyGlnI 245
QY 1077 TGCTGAGATGAGATTATGAAGAGCTGCGTGAAGCCTTTGAAGAAATGATGCTGCTGT 1136
|||
Db 245 nAlaIySgIuMetValIeuGlnIleuIleArgAsp----- 256
QY 1137 TAACACCCACTCCGATCTTCCAGACCGTGAACCCCATACACCAATTTGGCCCGTTCC 1196
|||
Db 257 -----GlnGlyGlyPheArgGlyVal-----ArgAsnGlyIyGly----- 268
QY 1197 GCATATCACTCTTATCCAGACAGAGATTTGATCTTTCAATCCCAACCAAGAGCTGT 1256
|||
Db 269 -----SerArgIleGlyIyAsnGlnIyIleAspValaProIleProArgPheAlaIy 286
QY 1257 GGGCGGCATCATCGGAGAAAGGGGACACATCAACAGCTGGGAGATTGGCGCGAGC 1316
|||
|||

```

Db      286  lglYlIvalIleGlYArGAnGlYlMeIlelYsIySlleGlInAsnAspAlaGlYVa 306
QY      1317  CTCTATCAAGATTGCCCCCTGCGAAGGCCAGCTCAGCAAGATGTCATCATCAC 1376
Db      306  lArgIleGlInPheYsProAspArgIly---ThrThrProIuArgIleAlaIleTh 325
QY      1377  CGGGCCACCGGAA---GCCAGTTCAGGCCAG----- 1407
Db      325  rGlYProProAspArgCysGlInHsAlaIleAlaIleleThrAspIleuAspSerVa 345
QY      1408  -----GGACGATCTTTGGGAAA----- 1425
Db      345  lGlInAlaIlyAsnProGlYglYProGlYProGlYglYArGlyArGlyArGlyInGl 365
QY      1426  -----CTGAAGAGGAAACTCTTTAACCCCAAGA 1457
Db      365  yAsnTrpAsnMetGlYProProGlYglYleuInGlInPheAsnPhelle----- 381
QY      1458  AGAAGTGAAGCTGGAAGCGCATATCAGAGTGCCTTCCTTCACAGCTGGCGGATTTGG 1517
Db      382  -----ValProThrGlYlYsThrGlYleuIleleGl 392
QY      1518  CAAAGTGGCAAGACCGTGAAGCAAGTGAAGTTCATTAACAGTGCAGAAAGTATGTCGCC 1577
Db      392  yLysGlYglYgluThrIleYsSerIleSerGlInInsErGlYAlaArgIleGlInleuGl 412
QY      1578  TCGTGAACCAAGCCAGATGAATAAGAGAGT---ATCGCAGAAATTATGGGGAC-- 1632
Db      412  nArgAsnProProAsnAlaAspProAsnMetLysLeuPheThrIleArgGlYThrPr 432
QY      1633  -----TCTTTGCTAGCCAGCTGCACAGCCGCAAGATCAGGGGAAATTGTACACA 1682
Db      432  oGlInGlInleAspTYAlaArgGlInleuIleGlInuYsIleGlYglYProValAsnPr 452
QY      1683  GGTGAAGCAGCAGAGCAGAAATACCCCTCAGGAGTGCCTCAGCGCAGCAAGTGAAG 1742
Db      452  oLeuGlYProProValProHisGlYProHisGlYVal-ProGlYProHisGlYProProG 472
QY      1743  CTCCTCAGGCGACCGCAAAACAGAGATGATGACCTTCCACACCTGACAGAAAG 1802
Db      472  lYProProGlY-ProGlY-----ThrPro-----Met 480
QY      1803  AGACCAAGCGACCGACCGCATCGGGAGCAACCAAGCATGTAGAGATGAGAAGT 1862
Db      481  GlYProTYAsnProAlaProTYAsnProGlYPro----- 492
QY      1863  CTCGCGAGCGCGCCAGGACTCTGCCGAGG---CCCTGAGAACCCGAGGGCCGAGGAGG 1919
Db      493  -----ProGlYProAlaProHisGlYProProAlaProTYAlaProGlInGlY 508
QY      1920  GGGCGGAGAGGTCAAGCGGTTGCCGAA-----CCACCGAGCCCGCGCC 1964
Db      509  TrpGlYAsnAlaTYrProHisTrpGlInGlInAlaProProAspProAla 525

```

RESULT 14
PCT-US94-01782-10

Sequence 10, Application PC/TUS9401782
GENERAL INFORMATION:
APPLICANT: THE GOVERNMENT OF THE UNITED STATES
APPLICANT: AS REPRESENTED BY THE SECRETARY, DEPARTMENT OF
APPLICANT: HEALTH AND HUMAN SERVICES
TITLE OF INVENTION: NOVEL FUSE BINDING
TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & PINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK

```

COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01782
FILING DATE: 22-FEB-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/021,608
FILING DATE: 22-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAM S. FETTER
REGISTRATION NUMBER: 26,728
REFERENCE/DOCKET NUMBER: 2026-4063PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 643
TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: Unknown
MOLECULE TYPE: Peptide/Protein
HYPOTHETICAL: No
ORIGINAL SOURCE: Human
ORGANISM: Human
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE: HL60
ORGANELLE:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PCT-US94-01782-10
Alignment Scores:
Pred. No.: 2.06e-14
Score: 262.00
Percent Similarity: 37.56%
Best Local Similarity: 23.37%
Query Match: 4.45%
Gaps: 25
US-09-270-437D-8 (1-3283) x PCT-US94-01782-10 (1-643)
QY      336  CCTCCCTCAGCTGAGTGGAGGTGTTGATGACCTTTGGCTCAATATGAGGACATGGA 395
Db      10  ProSerSerIySerAlaGlYglYglY-----GlYglYglY 22
QY      396  GAATGTGGAACAAGTCAACAACAACAAGAACCGCGCTTGTCAAGTCAATATGCAAC 455
Db      23  GlYglYglY-GlYValAsnAspAlaPheLysAspAlaLeuGlInArg-----AlaAr 39
QY      456  AAGAGAAAGCAAAATATGCCATGAGAGAAAGCTTAAGCGGGCATAGTTTGAAGACTCTC 515
Db      39  gInIleAlaAlaLysIleGlYglYAspAlaGlYThrSerLeuAsnSerAsnAspTYrGl 59
QY      516  CTTGAAGATTTCCATCAATCCCGATGAAGAGTGCCTCCCTTCGCCCTCAGCGAGC 575
Db      59  yTYrGlYglYglInLysArgProLeuGlInAspGlYAspGlInProAspAlaLysValAl 79
QY      576  CCAGCGTGGGAGCACTCTTCCCGGAGCAAGCCGCCCTGGGGGCACTTCTCAGGC 635
Db      79  aProGlInAsnAspSerPheGlYThrIn-----LeuProProweThsGlInGlInGl 96

```


ORGANISM: Human
CELL LINE: HL60
US-08-021-608D-8

Alignment Scores:

Pred. No.:	8.21e-14	Length:	590
Score:	255.00	Matches:	116
Percent Similarity:	38.52%	Conservative:	72
Best Local Similarity:	23.77%	Mismatches:	178
Query Match:	4.33%	Indels:	122
DB:	1	Gaps:	21

US-09-270-437D-8 (1-3283) x US-08-021-608D-8 (1-590)

```

QY GTCCCAACCCAGTTTGGTGGCCATCCGAAAGAGGGCTTGACATTAAGACATC 726
   |||||
Db ValProAspGlyMetValGlyPheIleIleGlyArgGlyGlnGlnIleSerArgIle 110
QY 727 ACTAAGCAGACCCAGTCCCGGTAGATTCATAGAAAAGAACTCTGGAGCTGCAG 786
   |||||
Db 111 GlnGlnIleSerGlyCysIleIleGlnIle--AlaProAspSerGlyIleuProGlu 129
QY 787 AAGCTGTCACCATCCATGCCAGAGGGGACTTCTGAAGCATGCCGATGATTCTT 846
   |||||
Db 130 ArgSerCysMetLeuThrGlyThrProIleuSerValGlnSerAlaIleArgIleuAsp 149
QY 847 GAATCATGACAGAAA-----GAGGCAGATGAGACCAACTA 882
   |||||
Db 150 GlnIleValGlnIleGlyAlaProAlaProGlyPheHisIleGlyAspGlyProGlyAsn 169
QY 883 GCCGAGAGAGATCTCTGAAATCTTGGACACACATGGCTTGGAAAGACTGATTGA 942
   |||||
Db 170 AlaValGlnIle-----GlnIleMetIleProIleAspIleGlyIleuValIleGly 186
QY 943 AAGAAGGAGAAATTGAAAGAAATTGAACATGAACAGGACCAAGATTAACAATTCTCA 1002
   |||||
Db 187 LysGlyGlyGlnThrIleLysGlnIleuGlnIleuArgIleValIleMetValMet--- 205
QY 1003 TCTTGGCAGGATTTGACATATACCAACCGGAAAGACC-----ATCACT 1047
   |||||
Db 206 ---IleGlnAsp-----GlyProGlnAsnThrGlyAlaAspLysProIleuArg 220
QY 1048 GTGAAGCGGACAGTTGAGCGCTGTGCCAGTGTGATAGATTAAGAAAGAGCGGT 1107
   |||||
Db 221 IleThrGlyAspProIleLysValGlnGlnIleAlaIleGlnMetValIleuGlnIleArg 240
QY 1108 GAGGCTTTGAAATGATATGCTGGCTTAAACCACTCCGATATCTTCCAGCCTG 1167
   |||||
Db 241 Asp-----GlnGlyGlyPheArgGlnVal 248
QY 1168 TACCCCATCACCAAGTTTGCCCGCTTCCGATCATCTTATCCAGACGAGATT 1227
   |||||
Db 249 ---ArgAsnGlnIleIle-----SerArgIleGlyGlnIleGlnIle 261
QY 1228 GTGATATCTTTCATCCCAACCAAGCTGTGGCGCCCATCATCGGAGAAAGAGGCGACAC 1287
   |||||
Db 262 IleAspValProIleProArgPheAlaValIleValIleGlyArgAsnGlyGlnMet 281
QY 1288 ATCAAAAGCTGGGAGATTCGCGGAGCTCTATCAAGATTGCCCTGCGGAAAGGCCCA 1347
   |||||
Db 282 IleLysLysIleGlnAsnAspAlaGlyValArgIleGlnPheLysProAspAspGly--- 300
QY 1348 GAGCTCAGCAGAAAGATGTCATCATCAACCGGCGCCACCGGAA---GCCAGTTCAAGGCC 1404
   |||||
Db 301 ThrThrProGlnArgIleAlaGlnIleThrGlyProProAspArgCysGlnHisAlaIle 320
QY 1405 CAG----- 1407
   |||||
Db 321 GlnIleIleThrAspLeuLeuArgSerValGlnAlaGlnAsnProGlyGlyProGlyPro 340
QY 1408 ---GAGCGGATCTTGGGAAA-----CTG 1428
   |||||
Db 341 GlyGlyArgGlyArgGlyArgGlyGlnGlyAsnTrpAsnMetGlyProProGlyGlyLeu 360

```

```

QY 1429 AAGAAGAAAATCTTCTTAACCCCAAGAAAGTGAAGCTGAAAGCCCATATCAGATG 1488
   |||||
Db 361 GlnGlnPheHisnHelle-----Val 367
QY 1489 CCTCTTCCACAGCTGCCCGGTGATTTGCAAGGTGGCAAGCCGAAAGCAATCGAG 1548
   |||||
Db 368 ProThrGlyLysThrGlyLeuIleIleGlyLysGlyGlnIleuTrpIleLysSerIleSer 387
QY 1549 AACTTAACGATGCGAAGTCACTGCTGCTGACCAAAAGCCAGATGAATAATGAGGAA 1608
   |||||
Db 388 GlnGlnSerGlyAlaArgIleGlnIleuGlnArgAsnProProProAsnAlaAspProAsn 407
QY 1609 GTG---ATCGCAGAAATTAATCGGCAC-----TTCTTGTCTAGCAAGATGCA 1653
   |||||
Db 408 MetLysIleuPheThrIleArgGlyThrProGlnGlnIleAspIleArgGlnIleu 427
QY 1654 CAGCGCAAGATCAGAGAAATTGTACACAGGTGAAGCAGCAGAGCAGAAATACCTCAG 1713
   |||||
Db 428 GlnGlnLysIleGlyGlyProValAsnProLeuGlyProProValProHisGlyProHis 447
QY 1714 GAGTCGCTTCACAGCGCAGCAAGTGAAGCTCCCAACAGCAGCAACAAACAGATGA 1773
   |||||
Db 448 GlyVal-ProGlyProHisGlyProProGlyProProGly-ProGly----- 462
QY 1774 ATGTAGCCCTTCCACACCTGACAGATGAGACCAAGCCAGCCAGATCGGAGCA 1833
   |||||
Db 463 ---ThrPro-----MetGlyProIleAsnProAlaProIleAsnProG 476
QY 1834 AACCAAAACCATCTGAGAAATGAGAAATGAGATCTGCGAGGCGGCAAGTCTGCGAGG 1892
   |||||
Db 476 LysPro-----ProGlyProAlaProHisG 484
QY 1893 ---CCCTGAACCCCGAGGCGCGAGAGGCGGAGAGGTGAGCGAGTTGCCAGAA- 1949
   |||||
Db 484 LysProProAlaProIleProGlnGlyTrpGlyAsnAlaTrpProHisIleTrpGlnGln 504
QY 1950 ---CCAGCGAGCCCGCC 1964
   |||||
Db 504 ImlAlaProProAspProAla 510

```

Search completed: July 23, 2004, 11:20:10
Job time : 65.5087 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 23, 2004, 11:15:41 ; Search time 156.647 Seconds

(Without alignments)
13126.119 Million cell updates/sec

Title: US-09-270-437D-8

Perfect score: 5886

Sequence: 1 ggcagcgagagagcgcagga.....aaccttgaaatgttattc 3283

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1288442 seqs, 313154207 residues

Total number of hits satisfying chosen parameters: 2576884

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing:

Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cg2_1/USPRO.spool/p/US09270437/runat_23072004.095742.27560/app.query.fasta_1.5562
-DB=Published Applications AA -QFMT=fastan -SUFFIX=rabp -MINMATCH=0.1
-LOOPC=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09270437@CGN 1.1.35 @runat_23072004.095742.27560
-NCPU=6 -ICPU=3 -NO MAP -LARGOQUERY -NEG SCORES=0 -WAIT -DSEBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:*

1: /cg2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cg2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cg2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cg2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cg2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cg2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cg2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cg2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cg2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cg2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
11: /cg2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cg2_6/ptodata/2/pubpaa/US09C_NEW_PUB.pep:*
13: /cg2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cg2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
15: /cg2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cg2_6/ptodata/2/pubpaa/US10C_NEW_PUB.pep:*
17: /cg2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cg2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	2838	48.2	556	14	US-10-097-340-147	Sequence 147, App
2	2838	48.2	556	16	US-10-648-553-182	Sequence 182, App
3	2808	47.7	620	9	US-09-764-864-1116	Sequence 1116, App
4	2745.5	46.6	587	15	US-10-313-986-501	Sequence 501, App
5	2584.5	43.9	555	15	US-10-262-445-40	Sequence 40, App1
6	1940	33.0	577	15	US-10-313-986-500	Sequence 500, App
7	1934	32.9	577	9	US-09-873-637-2	Sequence 2, App11
8	1886	32.0	589	15	US-10-313-986-486	Sequence 486, App
9	1875	31.9	579	9	US-09-735-705-348	Sequence 348, App
10	1875	31.9	579	9	US-09-850-716A-348	Sequence 348, App
11	1875	31.9	579	9	US-09-897-778-348	Sequence 348, App
12	1875	31.9	579	9	US-09-897-778-446	Sequence 446, App
13	1875	31.9	579	9	US-09-897-778-449	Sequence 449, App
14	1875	31.9	579	12	US-10-007-700-348	Sequence 348, App
15	1875	31.9	579	12	US-10-007-700-446	Sequence 446, App
16	1875	31.9	579	12	US-10-007-700-449	Sequence 449, App
17	1875	31.9	579	14	US-10-117-982-348	Sequence 348, App
18	1875	31.9	579	14	US-10-117-982-446	Sequence 446, App
19	1875	31.9	579	14	US-10-117-982-449	Sequence 449, App
20	1875	31.9	579	14	US-10-117-982-480	Sequence 480, App
21	1875	31.9	579	15	US-10-313-986-446	Sequence 446, App
22	1875	31.9	579	15	US-10-313-986-449	Sequence 449, App
23	1875	31.9	579	15	US-10-313-986-480	Sequence 480, App
24	1875	31.9	579	15	US-10-313-986-480	Sequence 480, App
25	1870	31.8	586	9	US-09-850-716A-427	Sequence 427, App
26	1870	31.8	586	12	US-10-007-700-427	Sequence 427, App
27	1870	31.8	586	12	US-10-007-700-427	Sequence 427, App
28	1870	31.8	586	14	US-10-117-982-427	Sequence 427, App
29	1870	31.8	586	15	US-10-313-986-427	Sequence 427, App
30	1868	31.7	579	9	US-09-735-705-176	Sequence 176, App
31	1868	31.7	579	9	US-09-850-716A-176	Sequence 176, App
32	1868	31.7	579	9	US-09-897-778-176	Sequence 176, App
33	1868	31.7	579	12	US-09-466-396A-176	Sequence 176, App
34	1868	31.7	579	12	US-10-007-700-176	Sequence 176, App
35	1868	31.7	579	14	US-10-117-982-176	Sequence 176, App
36	1868	31.7	579	15	US-10-313-986-176	Sequence 176, App
37	1865	31.7	579	14	US-10-117-982-484	Sequence 484, App
38	1865	31.7	579	15	US-10-313-986-484	Sequence 484, App
39	1326	22.5	422	16	US-10-408-765A-2088	Sequence 2088, App
40	812.5	13.8	261	9	US-09-764-864-1114	Sequence 1114, App
41	803	13.6	171	9	US-09-764-864-1119	Sequence 1119, App
42	782.5	12.3	250	9	US-09-764-864-1532	Sequence 1532, App
43	750	12.7	171	9	US-09-764-864-1536	Sequence 1536, App
44	676.5	11.5	192	9	US-09-764-864-1117	Sequence 1117, App
45	357	6.1	81	14	US-10-117-982-476	Sequence 476, App

ALIGNMENTS

RESULT 1
US-10-097-340-147
Sequence 147, Application US/10097340
Publication No. US20030087250A1
GENERAL INFORMATION:
APPLICANT: John MONAHAN
APPLICANT: Manjula GANNAVAPU
APPLICANT: Sebastian HOERSCH
APPLICANT: Shubhangi KAMATKAR
APPLICANT: Steve E. KOVARS
APPLICANT: Rachel E. MEYERS
APPLICANT: Michael MORRISSEY
APPLICANT: Peter OLANDT
APPLICANT: Ami SEN
APPLICANT: Peter VEIBY
APPLICANT: Gordon B. MILLS
APPLICANT: Robert C. BAST, Jr.
APPLICANT: Karen LU
APPLICANT: Rosemarie SCHMANT
APPLICANT: Xumei ZHAO
APPLICANT: Karen GLATT
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification, Assessment, Prevention, and Therapy of Ovarian Cancer
FILE REFERENCE: MRI-030

CURRENT APPLICATION NUMBER: US/10/097,340
CURRENT FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: 60/276,025
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/325,149
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/276,026
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/324,967
PRIOR FILING DATE: 2001/09/26
PRIOR APPLICATION NUMBER: 60/311,732
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: 60/325,102
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/323,580
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 363
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 147
LENGTH: 556
TYPE: PRT
ORGANISM: Homo sapiens
US-10-097-340-147
Alignment Scores:
Pred. No.: 4,296-221 Length: 556
Score: 2838.00 Matches: 556
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 48.22% Indels: 0
Gaps: 0
DB: 14
US-09-270-437d-8 (1-3283) x US-10-097-340-147 (1-556)
QY 70 ATGATGAACAGCTTATCATCCGGAACTGAGCCCGCGTCAACCGCCGACGACTCCGG 129
DB 1 MetMetAsnysleutyrlleeglyshenleuserProalaValThrAlaaspheleuarg 20
QY 130 CAGCTCTTTGGGACAGAACTGCCCCCTGGCGGACAGGTCTCTGTTAAAGTCCGGCTAC 189
DB 21 GlnleuPheGlyAspArglyLeuProleuAlaGlyGlnValleuLeuysSerglyTyr 40
QY 190 GCCTTGTGAGTACACCCCGACGAACTGGGGCATCCGGCCATCGAGACCTCTCGGGT 249
DB 41 AlaPheValAspTyrProaspGlnAsnThrPalaIleArgAlaIleGlnThrleuSergly 60
QY 250 AAGTGAATTGCATGGGAAATCATGAAAGTATTACTCACTCTTAAAGCTPAAG 309
DB 61 LysValGlnleuHisGlyIleMetGluValAspTyrServalSerylslsleuArg 80
QY 310 AGCAGGAAATTCAGATTGAAACATTCCTCTCTCACTGACGTGGAGGTGTTGGATGA 369
DB 81 SerArglyIleGlnIleArgAsnIleProIleHisleuGlnIlePrgIleValleuAspGly 100
QY 370 CTTTGGCTCAATATGGGACAGTGGAGATGTGGAACTGCAACAGACACAGAAACC 429
DB 101 LeuLeuAlaGlnTyrGlyThrValGlnAsnValGlnleuValAsnThrAspThrGlnThr 120
QY 430 GCGGTGTCAACGTCATATGCAACAGAGAAAGAAACAAATAGCCATGAGAACTTA 489
DB 121 AlaValAlaAsnValThrTyrAlaThrArgGlnGlnAlaIleAlaMetGlnIlyLeu 140
QY 490 AGCGGGCATCAGTTGAACTACTCTTCAAGTTTCTTCAATCCCGGATGAAAGAGTG 549
DB 141 SerGlyHisGlnPheGlnAsnTyrSerPheIlyIleSerylTyrIleProaspGlnIleVal 160
QY 550 AGCTCCCTTGGCCCTCAAGCGAGCCAGCGGTGGAGACCACTTCCCGGAGCAAGGC 609
DB 161 SerSerProserProProGlnIleArgAlaGlnIleArgIleAspHisSerSerArgGlnGlnGly 180
QY 610 CAGCCCTCTGGGGCACTTCTCAGGCCAGACAGATTGATTTCCGCTGCGGATCTGGTC 669
DB 181 HisAlaProGlyGlyThrSerGlnAlaArgGlnIleAspPheProleuArgIleLeuVal 200

QY 670 CCAACCAAGTTTGTGGTCCATCATCGGAAGAGGCGCTTGACCATTAAGAACATCACT 729
DB 201 ProThrGlnPheValGlyAlaIleIleGlyLysGlnIlyLeuThrIleLysAsnIleThr 220
QY 730 AAGCAGACCCAGTCCCGGTGATGATTCATGAAAAGAACTCTGAGCTCGCAGAGAG 789
DB 221 LysGlnThrGlnSerArgValAspIleHisArgLysGlnAsnSerglyAlaAlaGlnLys 240
QY 790 CCGTCAACCATCCATGCGCACCCAGAGGGAGCTTCTGAAGCATGCGCATGATCTTGA 849
DB 241 ProValThrIleHisAlaThrProGlnIlyThrSergIleAlaCysArgMetIleLeuGln 260
QY 850 ATCATGCAAGAAAGGCGAGATGAGCCAACTAGCCGAAAGATTCCTTGAAAATCTTG 909
DB 261 IleMetGlnLysGlnAlaAspGlnThrLysleuAlaGlnIleIleProleuLysIleLeu 280
QY 910 GCACCAATGGCTGTGGTGAAGACGTGATGAAAAGAAAGGAGAAATTTGAAAATTT 969
DB 281 AlaHisAsnGlyLeuValGlyArgleuIleGlyLysGlnGlyArgAsnleuLysIle 300
QY 970 GAACATGAACAGGGACCAAGATAACAATCTCATCTTTCAGAGATTGACATATACAAC 1029
DB 301 GlnHisGlnThrGlyThrLysIleThrIleSerSerleuGlnAspLeuSerIleTyrAsn 320
QY 1030 CCGAAGAAACCATCATCTGTGAAGGCAACAGTTGAGGCTTGCCAGTCTGAGATGAG 1089
DB 321 ProGlnArgThrIleThrValLysGlyThrValGlnAlaCysAlaSerAlaGlnIleGln 340
QY 1090 ATTATGAAGAAAGCTCGCGTGAAGCCCTTGAAATGATGCTGCGTGTAAACCCACTCC 1149
DB 341 IleMetLysLysLeuArgGlnAlaPheGlnAsnAspMetLeuAlaValAsnThrHisSer 360
QY 1150 GGATCTCTCCAGCTGTACCCCATCAACAGTTTGCGCGCTCCCGCATCATCACTCT 1209
DB 361 GlyTyrPheSerSerleutyrlProHisGlnPheGlyProPheProHisIleHisSer 380
QY 1210 TATCCAGACAGAGATTGTGAATCTTCAATCCCAACCGAGCTGTGGCGCATCAATC 1269
DB 381 TyrProGlnGlnGlnIleValAsnleuPheIleProThrGlnAlaValGlyAlaIleIle 400
QY 1270 GGGAGAAAGGGGCGACATCAACAGCTGGGAGATTGGCGGACCTCTTCAAGATT 1329
DB 401 GlyLysLysGlyAlaHisIleLysGlnleuAlaArgPheAlaGlyAlaSerIleLysIle 420
QY 1330 GCGCCCTGCGGAGCGCCAGCATGACGTCAGGAAAGAGATGTCATCATCAACCGGCA 1389
DB 421 AlaProAlaGlnIlyProaspValSergIleArgMetValIleIleThrGlyProProGln 440
QY 1390 GCCCAGTTCAAGGCCCGAGGACGGAATCTTTGGAAACTGAAAAGAGAAACTCTTAA 1449
DB 441 AlaGlnPheLysAlaGlnIlyArgIlePheGlyLysleuLysGlnGlnAsnPheAsn 460
QY 1450 CCAAAAGAAAGATGAAGCTGGAAGGCGATATCAGAGTGCCCTTTCACAGCTGCGCG 1509
DB 461 ProLysGlnIleValLysleuGlnAlaHisIleArgValProSerSerThrAlaGlyArg 480
QY 1510 GTGATTGGCAAGAGTGGCAAGACCGTGAACGAACGCAACTTAACAGATGCAAGATC 1569
DB 481 ValIleGlyLysGlyGlyLysThrValAsnGlnleuGlnAsnleuThrServalAlaGlnVal 500
QY 1570 ATGTGCTGTGATCAACACGCGCATGAAATGAGAAATGAGATGATCCGCAAAATTA 1629
DB 501 IleValProArgAspGlnThrProaspGlnAsnGlnIleValIleValAlaGlyIleIleGly 520
QY 1630 CACTTCTTCTAGCCGACATGCAACAGCGCAAGATTCAGGGAATTTGACAAACAGTGA 1689
DB 521 HisPhePheAlaSerGlnThrAlaGlnIleArgLysIleArgGlnIleValGlnGlnValLys 540
QY 1690 CAGCAGAGCAGAAATACCTCAGAGAGTGCCTCACAAGCGCAGCAAG 1737
DB 541 GlnGlnGlnGlnLysTyrProGlnIlyValAlaSerGlnIleArgSerylLys 556

RESULT 2
US-10-648-593-182
; Sequence 182, Application US/10648593
; Publication No. US20040106132A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF GENES FOR PREDICTING ACTIVITY OF COMPOUNDS THAT
; TITLE OF INVENTION: INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE KINASES AND/OR
; FILE REFERENCE: D0273 NP
; CURRENT APPLICATION NUMBER: US/10/648,593
; PRIOR FILING DATE: 2003-08-26
; PRIOR FILING NUMBER: 60/406,385
; NUMBER OF SEQ ID NOS: 557
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 182
; LENGTH: 556
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-648-593-182

Alignment Scores:
Pred. No.: 4.29e-221 Length: 556
Score: 2838.00 Matches: 556
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 48.22% Indels: 0
DB: 16 Gaps: 0
US-09-270-437d-8 (1-3283) x US-10-648-593-182 (1-556)

QY 70 ATGATGAACAAGCTTACATGAGGACCTGAGCCCGCCGCTCACGCCGAGACGACTCTCCG 129
Db 1 Methectsnhslyseutyrlleghlyashnleuserproalvalthralspaspdeuarg 20
QY 130 CAGCTCTTTGGGAGCAGAGAAAGCTGCCCCGCGGAGACAGGCTCTGTAAGTCCGCGTAC 189
Db 21 GlnleuheghlysparglyseuProleuhalaglyglnvalleuLeuylsSerGlyTyr 40
QY 190 GCCTTGCTGACATACCCGACACAGAACTGGGCCATCCGCCCATGAGACCTCTCGGAGT 249
Db 41 AlalphevalaspyrProaspGlnasntrpAlalleargAlalleghutnrLeuSerGly 60
QY 250 AAGGTGAATTCATGAGGAAATCATGAGATTGATTACTCAGCTCTTAAGCTTAAG 309
Db 61 LysValGlnleuhsisglylysilemeGlnuValaspyrSerValSerLysLysLeuArg 80
QY 310 AGCAGAAATTCAGATTGAAAATCCCTCTCACCCTGCAAGTGGAGGTGTTGGATGA 369
Db 81 SerArglyseileglnlleargasnilleProProHisleuGlntrpGlnuValleuaspGly 100
QY 370 CTTTGGCTCAATATGGGAGACAGTGGAGAAATGGAACATGCAACACAGACAGAAACC 429
Db 101 LeuLeuhalaglnlyrGlnthrValGlnasnValGlnGlnValasnThrAspThrGlnThr 120
QY 430 GCCGTGTCAACCTCATATGCAACAGAGAGAGAAATAGCCATGAGAGAGCTA 489
Db 121 AlalValalasnValthrThrAlatnrAspGlnuValalysillealasnGlnLysLeu 140
QY 490 AGCGGCGATCAGTTGAGAACTACTCTTCACAGATTCTTACATCCCGATGAGAGAGT 549
Db 141 SerGlyHisGlnpheglnasnlyrSerPheLysileSerYrilleProaspGlnGlnuVal 160
QY 550 AGCTCCCTTCGCCCCCTCAGCGAGCCGACGCGTGGGAGCCTCTTCCGCGAGAGAGC 609
Db 161 SerSerProserProProGlnArgAlaglnArgLysAspHisSerSerArgGlnGlnGly 180
QY 610 CAGGCCCTCGGGGAGCTTCTCAGCGCAGACAGATTGATTTCCCGTGGGAGCTCGGTC 669
Db 181 HisAlaProGlyGlyHisSerGlnAlaGlnLysAspPheProleuArgLysLeuVal 200
QY 670 CCACCCCAAGTTTGTGTGTCATCATGAGAAAGAGGCGCTTGACATAAGACATCACT 729

Db 201 ProthrGlnpheValGlyAlallelleghlyLysGlnGlyLeuThrilleLysasnillethr 220
QY 730 AAGCAGACCCAGTCCCGGGTAGATATCCATGAAAAAGAACTCGAGCTCCAGAGAG 789
Db 221 LysGlnThrGlnSerArgValasprillehsarglysglnasnSerGlyAlalaglnLys 240
QY 790 CCTGACCATTCATGCCACCCAGAGGGGACTTGAAGCATGCCGATGATTTCTGGA 849
Db 241 ProValThrillehsAlatnrProGlnGlythrSerGlnuValaCysArgMetilleuGln 260
QY 850 ATCATGCAAGAAAGGCGATGAGACCAACTAGCCGAAGACATTCCTGAAAATCTTG 909
Db 261 llemeGlnLysGlnuAlaaspGlnthrLysLeuAlaglnGlnilleProleuLysilleu 280
QY 910 GCACACATGCGCTTGTGTTGAAGACTGATTGAAAAAGAGCAGAAATTTGAAGAAAT 969
Db 281 AlahlsasnGlyLeuValGlyArgleuilleghlysglnGlyAsnleuLysLysile 300
QY 970 GAACATGAACAGGACCCAGATACAAATCTCATCTTTGACAGATTGACATATACAC 1029
Db 301 GlnhsGlnThrGlyThrLysilleThrilleSerSerleuGlnaspLeuSerilleLysn 320
QY 1030 CCGAAGAAACCATCAGCTGTAAGGCGACACTTGAAGCTTGTCCAGTGTGATAGTAG 1089
Db 321 ProGlntrGlnThrilleThrValLysGlyThrValGlnuAlaCysAlaseralaglnilleGlu 340
QY 1090 ATTATGAAGAAAGCTGCGTGAAGGCTTTGAAAAATGATATGCTGCGCTGTTAAACCCACCTC 1149
Db 341 llemeLysLysLeuAlaagglnAlapheGlnasnspMetleuAlalasnThrHisSer 360
QY 1150 GAATATCTTCTCAGCTGTACCCCATCACCAATTTGACCGCTCCGCATCATCACTCT 1209
Db 361 GlyTyrPheSerSerleutyProHisGlnpheGlyProPheProHisHisSer 380
QY 1210 TATCAGAGAGAGAGATTGGAATCTCTTATCCCAACCCAGAGCTGTGGGCGCATATC 1269
Db 381 TyrProGlnGlnGlnilleValasnLeupheilleProthrGlnAlaValalGlyAlallelle 400
QY 1270 GGGAGAGAGGGGAGCAGATCAAAACAGCTGGCGAGATTGGCGGAGGCTCATAGATT 1329
Db 401 GlyLysLyslelyAlahlsilleLysGlnleuAlahrpheAlaglyAlasertilleLysile 420
QY 1330 GCCCTTCGCGAGGCGCCAGAGCTGACGGAAGAGATGTCATCAACCGGCGCACCGGAA 1389
Db 421 AlaprolalaglnGlyProaspValSerGlnuArgMetValillellethrGlyProProGln 440
QY 1390 GCCCAGTTCAAGGCGCAGAGGAGCGATCTTTGGGAACTGAAAGAGAAACTTCTTTAAC 1449
Db 441 AlaglnPheLysAlaglnGlyArgillePheGlyLysLeuLysGlnGlnuasnPheasn 460
QY 1450 CCCAAAGAAAGAGTGAAGCTGGAAGCGCATATCAGAGTGCCTTTCACAGCTGGCGG 1509
Db 461 ProLysGlnGlnuValLysleuGlnuAlahlsilleArgValProserSerThrAlaglyArg 480
QY 1510 GTGATTGGCAAGAGTGGCAAGACCGTGAAGCAACTGCAAGAACTTAACCAAGTGAAGTC 1569
Db 481 ValilleGlyLyslelyGlyLysThrValasnGlnleuGlnasnleuThrSerAlaglnuVal 500
QY 1570 ATGTGCGCTCGTGAACCAAGCGCAGATGAAAAAGAGAGAGTGTGTCAGAAATTAATCGG 1629
Db 501 lleValProArgaspGlnthrProaspGlnasnGlnuValilleValalArgilleleGly 520
QY 1630 CACTTCTTTGCTAGCCAGACTGACAGCGCAAGATCAGGGGAAATTTATCAACAGGTAAG 1689
Db 521 HisPhePheAlasertlnthrAlaglnArgLysilleArgGlnilleValalGlnGlnuLys 540
QY 1690 CAGCAGAGACAGAAATACCTTCAAGGAGTCCCTCAACCGCAGCAAG 1737
Db 541 GlnGlnGlnGlnLysLysrProGlnGlyValalaserGlnArgSerLys 556
RESULT 3
US-09-764-864-1116

```
; Sequence 1116, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PAM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.10
; SEQ ID NO 1116
; LENGTH: 620
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (533)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1116

Alignment Scores:
Pred. No.: 1,226-218 Length: 620
Score: 2808.00 Matches: 560
Percent Similarity: 90.78% Conservatve: 1
Best Local Similarity: 90.61% Mismatches: 11
Query Match: 47.71% Indels: 46
DB: Gaps: 2

US-09-270-437D-8 (1-3283) x US-09-764-864-1116 (1-620)
QY 22 CGCCGGGTCACCGGGCCGGGAGCCGGGCTCTCGGGGAGAGACG-----ANG 72
DB 3 ArgArgTyrAlaCysArgTyrArgSerGlyIleProGlySerThrHisAlaSerGlyMet 22
QY 73 ATGAAGAAGCTTTACATCGGGAACTGAGCGCCGCGTCAACCGCCGACGCTCCGGCAG 132
DB 23 MetAsnLysLeuTyrIleGlyAsnLeuSerProAlaValThrAlaAspLeuArgLys 42
QY 133 CTCTTTGGGAGACAGAAAGCTGCCCTGGCGGACAGGTCCTGCTGAAGTCGGCTACGCC 192
DB 43 LeuPheGlyAspArgLysLeuProLeuAlaGlyGlnValLeuLeuLysSerGlyTyrAla 62
QY 193 TTCTGGACTACCCCGACCAAGACTGGGCTATCCGGGCCATCGAGACCCCTCGGGTAA 252
DB 63 PheValAspTyrProAspGlnAsnTyrAlaIleArgAlaIleGlnThrLeuSerGlyLys 82
QY 253 GTGGAAATTGCAAGGAAATCATGGAAGTTGATTACTAGTCTCTAATAAGCTPAAGAGC 312
DB 83 ValGlnLeuHisSerGlyLysIleMetGlnValAspTyrSerValSerLysLysLeuArgSer 102
QY 313 AGGAAATTCAGATTGCAAAATCCCTCTCCTACCTGACGTGGAGAGTGTTGATGACTT 372
DB 103 ArgLysIleGlnIleArgAsnIleProProHisLeuGlnTyrGlnValLeuAspGlyLeu 122
QY 373 TTGGCTCAATATGGAGACGTGGAATGTGGAACAATCAACAAGACAGAAACCGCC 432
DB 123 LeuAlaGlnTyrGlyThrValGlnAsnValGlnGlnValAsnThrAspThrGlnThrAla 142
QY 433 GTTGTCAGCTCAATATGCAACAAGAGAAAGCAAAATATAGCCATGAGAGAGCTAAGC 492
DB 143 ValValAsnValThrTyrAlaThrArgGlnGlnAlaLysIleAlaMetGlnLysLeuSer 162
QY 493 GGGGATCAGTTTGAACTACTCTCTTCAAGATTCTTCAATCCCGGATGAAGAGTGAGC 552
DB 163 GlyHisGlnPheGlnAsnTyrSerPheLysIleSerTyrIleProAspGlnGlnValSer 182
QY 553 TCCCTTCGCCCTTCACGACGCCCAAGCGGGGACCACTCTTCCCGGAGCAAGGCAC 612
DB 183 SerProSerProGlnArgAlaGlnArgGlyAspHisSerSerArgGlnGlnGlyHis 202
QY 613 GCCCTGGGGCACTTCTCAGGCGACAGATGATTCCCGCTGCGGATCCTGCTCC 672
```

```
DB 203 AlaProGlyGlyThrSerGlnAlaArgGlnIleAspPheProLeuArgIleLeuValPro 222
QY 673 ACCGATTTGTTGGTCATCATCGAAAGAGGGCTTGACCATTAAGACATCACTAAG 732
DB 223 ThrGlnPheValGlyAlaIleIleGlyLysGlnLysLeuThrIleLysAsnIleThrLys 242
QY 733 CAGACCCAGTCCCGGGTAGATATCCATTAAGAAAGAACTCTGGAGCTCGCAGAGAGGCT 792
DB 243 GlnThrGlnSerArgValAspIleHisArgLysGlnLysSerGlyAlaAlaGlnLysPro 262
QY 793 GTACCATTCATGACCCCGACAGGGGACCTCTGAGCATGCCGATGATCTTGAAATC 852
DB 263 ValThrIleHisAlaThrProGlnGlyThrSerGlnAlaCysArgMetIleLeuIle 282
QY 853 ATGCAGAAAGAGCAGATGAGACCAACTAGCCGAGAGATTCCTGTAATAATTGGCA 912
DB 283 MetGlnLysGlnAlaAspGlnThrLysLeuAlaGlnGlnIleProLeuLysIleLeuAla 302
QY 913 CACAAATGGCTTGTGGAAAGCTGATTTGGAAAAGAGGACAGAAATTTGAAGAAATTGAA 972
DB 303 HisAsnGlyLeuValGlyArgLeuIleGlyLysGlnGlyArgAsnLeuLysIleGln 322
QY 973 CATGAAGACAGGACCAAGATTAACAATCTCATCTTGCAGAGATTGAGCATATACACCG 1032
DB 323 HisGlnThrGlyThrLysIleThrIleSerSerLeuGlnAspLeuSerIleTyrAsnPro 342
QY 1033 GAAAGAACCATCATCTGTGAAGGGCAAGTTGAAGCCTTGCCAGTGTGATGATGATTT 1092
DB 343 GlnArgThrIleThrValLysGlyThrValGlnAlaCysAlaSerAlaGlnIleGlnIle 362
QY 1093 ATGAAGAAGCTCGGAGGCTTTGAAAATGATATAGCTGGCTGTATAC----- 1140
DB 363 MetLysLysLeuArgGlnAlaPheGlnAsnAspMetLeuAlaValAsnGlnAlaAsn 382
QY 1140 ----- 1140
DB 383 LeuIleProGlyLeuAsnLeuSerAlaLeuGlyIlePheSerThrGlyLeuSerValLeu 402
QY 1141 -----ACC 1143
DB 403 SerProProAlaGlyProArgGlyAlaProProAlaAlaProTyrHisProPheThrThr 422
QY 1144 CACTCCGATACTTCTCCAGCCTGTACCCCATCACAGATTGGCCCGTCCCGCATCAT 1203
DB 423 HisSerGlyTyrPheSerSerLeuTyrProHisHisGlnPheGlyProPheProHis 442
QY 1204 CACTCTTATCCAGAGAGAGGATTTGAACTCTGTATCCCAACCAAGGCTGAGGCGCC 1263
DB 443 HisSerTyrProGlnGlnGlnIleValAsnLeuPheIleProThrGlnAlaValGlyAla 462
QY 1264 ATCATCGGAGAAAGGGGGCACATCAACATCAACAGCTGGCGAGATTCCCGGAGCTCTATC 1323
DB 463 IleIleGlyLysLysGlyAlaHisIleLysGlnLeuAlaArgPheAlaGlyAlaSerIle 482
QY 1324 AAGATTGCCCTTCGGAAGGCCCAAGACGTGAGGAAGATGTGCATCATCACCGGSCCA 1383
DB 483 LysIleAlaProAlaGlnGlyProAspValSerGlnArgMetValIleIleThrGlyPro 502
QY 1384 CCGGAGCCAGATTCAAGGCCCAAGGACGAGATCTTTGGAACTGAAAGAGAAACTTC 1443
DB 503 ProGlnAlaGlnPheLysAlaGlnGlyArgIlePheGlyLysLeuLysGlnGlnAsnPhe 522
QY 1444 TTTAACCCCAAGAAAGAGTGAAGCTGGAAGCGCATATCAAGAGTCCCTCTTCCACAGCT 1503
DB 523 PheAsnProLysGlnGlnValLysLeuGlu**HisIleArgValProSerSerThrAla 542
QY 1504 GCGCGGCTGATTGGCAAGGTGCGCAAGCGTGAACGATGCGAGACTTAACAGTGCA 1563
DB 543 GlyArgValIleGlyLysGlyLysGlyLysThrValAsnGlnLeuGlnAsnLeuThrSerAla 562
QY 1564 GAAATCATCGGCTCGGACCAAGCCAGATGAAATGAGAAATGATGATGTCGTAATT 1623
DB 563 GlnValIleValProArgAspGlnThrProAspGlnAsnGlnGlnValIleValArgIle 582
```

Qy	1624	ATCGGGACACTCTTTTGTCTAGCCGACTGCACAGCCGCAATCATCGGGAATTTGTACACAG	1683
Db	583	lleglyshpshphealaserglntrhralaiglnarglysilargglullevalsgln	602
Qy	1684	GTGAGACGACGAGGAGCAATACCTCTCAGGGAGTCGCTCAACGACGACGACG	1737
Db	603	Vallyelnglnsglnuglnytrpoglnglyvalalaserglnarserlys	620
RESULT 4			
US-10-313-986-501			
/ Sequence 501, Application US/10313986			
/ Publication No. US20030236209A1			
GENERAL INFORMATION:			
/ APPLICANT: Foy, Teresa M.			
/ APPLICANT: McNabb, Andria			
/ APPLICANT: Matanabe, Yoshihiko			
/ APPLICANT: Reed, Steven G.			
/ APPLICANT: Wang, Tonglong			
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY			
/ TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER			
/ FILE REFERENCE: 210121.455C19			
/ CURRENT APPLICATION NUMBER: US/10/313, 986			
/ NUMBER OF SEQ ID NOS: 560			
/ SOFTWARE: FastSeq for Windows Version 4.0			
/ SEQ ID NO 501			
/ LENGTH: 587			
/ TYPE: PRT			
/ ORGANISM: Homo sapiens			
US-10-313-986-501			
Alignment Scores:			
Pred. No.: 1,41e-213			
Score: 2745.50			
Percent Similarity: 92.67%			
Best Local Similarity: 92.67%			
Query Match: 46.64%			
DB: 15			
Gaps: 1			
US-09-270-437D-8 (1-3283) x US-10-313-986-501 (1-587)			
Qy	73	ATGACAAGACTTTTACATCGGGGAACCTGAGCCCGCGCTCAACCGCCGACGACCTCGCGGAG	132
Db	1	MetasnlyslseuYrllleglyashneuserProAlaValhralaaspaspLeuArgGln	20
Qy	133	CTCTTTGGGGACAGAAAGCTGCCCCGCGCGGACAGGCTCGTGTGAAGTCCGGCTACGCC	192
Db	21	LeuphegllysparglyslseuProLeuAlaGlyGlnValLeuLeuylsserGlyTyrala	40
Qy	193	TTTCGGAAGTAAACCCGACCGAAGACTGGGCCATCCGGCCATCGAGACCTCTCGGGTAA	252
Db	41	PheValAspYrProAspGlnAsnTrpAlaIleArgAlaIleGlnThrLeuSerGlylys	60
Qy	253	GTGGAATTGCATGGGAANAATCATGGAAGTTACTCGATCGTCTTAAAGATAAGGAC	312
Db	61	ValGlnLeuHsiGlylysIleMetGlnValAspTySerValSerlyslseuArgSer	80
Qy	313	AGGAAATTTGAGATTGGAACATCCCTCTGACCTTCAGCTTGAGAGGTTGTGATGACTT	372
Db	81	ArglyslleglnIleArgAsnIleProProHsiLeuGlnTrpGlnValLeuAspGlyLeu	100
Qy	373	TTGGCTCAATATGGGACAGTGGGAATGTGGAAACAGTCAACACGACACGAAACGCC	432
Db	101	LeuAlaGlnIlyrGlyThrValGlnAsnValGlnGlnValAsnThrAspThrGlnThrAla	120
Qy	433	GTGTGACGCTCAATATGCAACACAGAGAAGCAAAATAGCCATGAGAGGTTAGC	492
Db	121	ValValAsnValThrTyralaThrArgGlnGlnIleAlaIleAlaMetGlnIlylser	140
Qy	493	GGGATACAGTTTGAAGACTACTCTTCAAGATTTTCTACATCCGGATGAAGAGTGAGC	552
Db	141	GlyHsiGlnPheGlnAsnTySerPheylsIleSerTyIleProAspGlnGlnValSer	160

QY	553	TCGCCCTTCGCCCTTCAGCGAGCCACGTCGTGGGAGCACCTCTCCCGGAGCAAGGCGAC	6112
Db	161	SerProSerProProGlnArgAlaGlnArgGlyAspHisSerSerHisArgGlnGlnHis	1805
QY	613	GCCCCCTGGGGCACTTTCTCAGGCGCAGACAGATTGATTTCCCGCTCGCATCTGGTCCC	6721
Db	181	AlaProGlyGlyThrSerGlnAlaArgGlnIleAspHeProLeuArgIleLeuValPro	2005
QY	673	ACCAGATTGTGGTGGCCATCATCGGAAAGAAGGGCTTGACCATTAAGAACATCATAAG	7325
Db	201	ThrGlnPheValGlyAlaIleIleGlyLysGlnGlyLeuThrIleLysAsnIleThrLys	2205
QY	733	CAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGAGCTGGCAGAGAGCCCT	7925
Db	221	GlnThrGlnSerArgValAspIleHisArgLysGlnSerSerGlyAlaAlaGlnLysPro	2405
QY	793	GTCACCATTCATGCCACCCCAAGAGGGGACTTCTGAAGCATGCCGATGATTTCTTGAATC	8525
Db	241	ValThrIleHisAlaThrProGlnGlyThrSerGlnAlaCysArgMetIleLeuGlnIle	2605
QY	853	ATCCAGAAAAGGCGACAGTGTAGACCAAACTAACCCGAGAGATTTCTTGAAATCTTGGCA	9125
Db	261	MetGlnLysGlnAlaAspGlnThrLysLeuAlaGlnGlnIleProLeuLysIleLeuAla	2805
QY	913	CACAAATGCGTGGTTGGAAAGACTGATTTGGAAAAAGAGCAGAAATTTGAAGAAATTTGA	9725
Db	281	HisAsnGlyLeuValGlyArgLeuIleGlyLysGlnGlyArgAsnLeuLysIleGln	3005
QY	973	CATGAACACGGGACCAAGATTAACATCTCATCTTTGGCAGATTGGCATATACAACCG	1035
Db	301	HisGlnThrGlyThrLysIleThrIleSerSerLeuGlnAspLeuSerIleLysAsnPro	3205
QY	1033	GAAGAAGCATCATCTGTGAAGGGCACAGTTGAGGCGCTGGCCAGAGCTGAGATGAGATT	1095
Db	321	GlnArgThrIleThrValLysGlnThrValGlnAlaCysAlaSerAlaGlnIleGlnIle	3405
QY	1093	ATGAAGAACCTCGTAGGCTCTTTGAAATGATATGATCGTGGCTGTAAAC-----	1145
Db	341	MetLysLysLeuArgGlnAlaPheGlnAsnAspMetLeuAlaValAsnGlnGlnAlaAsn	3605
QY	1140	-----	1145
Db	361	LeuIleProGlyLeuAsnLeuSerAlaLeuGlyIlePheSerThrGlyLeuSerValLeu	3805
QY	1141	-----ACC 1143	
Db	381	SerProProAlaGlyProArgGlyAlaAlaProProAlaAlaProTyHisProPheThrThr	4005
QY	1144	CACCTCCGAGATCTTCTCCAGCCTGTACCCCATCACAGTTTGGCCGCTGCCGATCAT	1205
Db	401	HisSerGlyTyPheSerSerLeuTyProHisHisGlnPheGlyProPheProHisHis	4205
QY	1204	CACCTCTTCCAGAGCAGAGATTTGTGAATCTTTCATCCCAACCAAGCTGTGGCGCC	12635
Db	421	HisSerTyProGlnGlnGlnIleValAlaAsnLeuPheIleProThrGlnAlaValGlyAla	4405
QY	1264	ATCATCCGGAAGAGGGGGCACAATCAATAAAGATGGCGAGATTCGCCGGAGCCTCATC	13235
Db	441	IleIleGlyLysLysGlyAlaHisIleLysGlnLeuAlaArgPheAlaGlyAlaSerIle	4605
QY	1324	AAGATTGCCCTCGCGAAGGCCCGACGTCAGCAAAAGATGTGCATCATCACCGGGCCA	13835
Db	461	LysIleAlaProAlaGlnGlyProAspValSerGlnArgMetValIleIleThrGlyPro	4805
QY	1384	CCGGAAGCCGATTCAGAGGCCCGCAGGACGGATCTTTGGAAACTGAAGAGAAACTTC	14435
Db	481	ProGlnAlaGlnPheLysAlaGlnGlyArgIlePheGlyLysLeuLysGlnGlnAsnPhe	5005
QY	1444	TTTAAACCCAAAGAGAGTGAAGCTGGAAGCGGATTCAGAGAGCCCTCTCCACAGCT	15035
Db	501	PheAsnProLysGlnGlnValLysIleGlnGlnAlaHisIleArgValProSerSerThrAla	5205

QY 1504 GCGCGGCTGATGTGCAAGGTGGCAAGACCGTGAACGAACTTGCAACTTACCGAGTGA 1563
Db 521 G1YArgValIleGlyLysGlyLysThrValAsnGluLeuGlnAsnLeuThrSerAla 540
QY 1564 GAAGTCATCGCGCTGCTGGACCAAGCCAGATGAAATGAGAAATGATCGTCAAAATT 1623
Db 541 GlnValIleValProGlnAspGlnThrProAspGlnAsnGluValIleValArgIle 560
QY 1624 ATCGGAGCACTTCTTGTGAGCCAGACTGACAGCGCAAGATCGAGAAATGTTCACACG 1683
Db 561 IleGlyHisPhePheAlaSerGlnThrAlaGlnArgLysIleArgGluIleValGlnGln 580
QY 1684 GTGAAGCAGCAGCAGCAGCAAAA 1704
Db 581 ValLysGlnGlnGlnGlnLys 587
RESULT 5
US-10-262-445-40
Sequence 40, Application US/10262445
Publication No. US20040014058A1
GENERAL INFORMATION:
APPLICANT: Alsobrook II, John
APPLICANT: Burgess, Catherine
APPLICANT: Catereron, Elina
APPLICANT: Chant, John
APPLICANT: Chaudhuri, Amitabha
APPLICANT: Edinger, Shlomit
APPLICANT: Gerlach, Valerie
APPLICANT: Giot, Loic
APPLICANT: Gorman, Linda
APPLICANT: Guo, Xiaojia
APPLICANT: Kekuda, Ramesh
APPLICANT: Mezes, Peter
APPLICANT: Millet, Isabelle
APPLICANT: Ooi, Chean Eng
APPLICANT: Paturajan, Meera
APPLICANT: Rieger, Daniel
APPLICANT: Spytek, Kimberly
APPLICANT: Taupier Jr., Raymond J.
APPLICANT: Zerhusen, Bryan
APPLICANT: Zhong, Haihong
APPLICANT: Zhong, Mei
TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS
FILE REFERENCE: 21402-462D
CURRENT APPLICATION NUMBER: US/10/262,445
CURRENT FILING DATE: 2002-10-01
PRIOR APPLICATION NUMBER: 60/327,454
PRIOR FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 60/327,917
PRIOR FILING DATE: 2001-10-09
PRIOR APPLICATION NUMBER: 60/328,029
PRIOR FILING DATE: 2001-10-09
PRIOR APPLICATION NUMBER: 60/328,056
PRIOR FILING DATE: 2001-10-09
PRIOR APPLICATION NUMBER: 60/328,849
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: 60/329,414
PRIOR FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 60/330,142
PRIOR FILING DATE: 2001-10-17
PRIOR APPLICATION NUMBER: 60/341,058
PRIOR FILING DATE: 2001-10-22
PRIOR APPLICATION NUMBER: 60/343,629
PRIOR FILING DATE: 2001-10-24
PRIOR APPLICATION NUMBER: 60/349,575
PRIOR FILING DATE: 2001-10-29
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 133
SOFTWARE: Curaseq1 version 0.1
SEQ ID NO 40
LENGTH: 555
TYPE: PRT

ORGANISM: Homo sapiens
US-10-262-445-40
Alignment Scores:
Pred. No.: 1,61e-200
Score: 2584.50
Percent Similarity: 94.60%
Best Local Similarity: 91.19%
Query Match: 43.91%
DB: 15
Gaps: 1
US-09-270-437d-8 (1-3283) x US-10-262-445-40 (1-555)
QY 70 ATGATGAAACAAGCTTTATCATGCGGAACTGAGCCCGCCGCTCAGCCGAGACCTCGG 129
Db 1 MetMetAsnLysLeuPheIleGlyAsnLeuSerProAlaValAlaGluAspLeuArg 20
QY 130 CAGCTCTTTGGGAGCAGGAAAGCTGCGCGGAGGACAGTCTGCTGAAGTCCGCTAC 189
Db 21 GlnLeuPheGlyAspArgLysLeuProLeuAlaGlyGlnValLeuLysSerArgTyr 40
QY 190 GCCTTCGTGACTACCCGACCCAGAACTGGGCGCATCGCGCCATGAGACCTCTCGGCT 249
Db 41 AlaPheValAspTyrProAspGlnSerTrpAlaIleArgThrIleGluThrLeuSerGly 60
QY 250 AAAGTGAATTGATGAGAAATATCATGAGATTTACTGATGCTCTTAAAGCTAAAG 309
Db 61 GlnValGluLeuHisGlyLysIleMetGluValAspTyrSerValSerIleLysLeuArg 80
QY 310 AGCAGGAAATTCAGATTGCAAAACATCCCTCTCACTGCGATGGAGGAGTGTGATGA 369
Db 81 SerArgAsnIleProIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspGly 100
QY 370 CTTTGGCTCATATATGAGACAGTGAAGATGTGAACAGTCAACACACACACAGAAAC 429
Db 101 LeuLeuAlaGlnTyrGlyThrValGluAsnValGluGlnValAsnThrAspThrGluThr 120
QY 430 GCCGTTGTCACGTCAATATGCAACACAGAAAGAAAGCAAAATAGCCATGGAGAGCTA 489
Db 121 AlaValAlaAsnValThrTyrAlaThrLysGluGluValLysIleAlaMetLysLysLeu 140
QY 490 AGCGGCGCATGATTGAGAACTACTCTTCAAGATTCTTACATCCCGATGAAGAGGTG 549
Db 141 SerGlyHisGlnPheGluAsnHisTyrPheLysIleSerTyrIleProAspAspGluVal 160
QY 550 AGCTCCCTTGGCCCTCAGCGAGCCAGACGCTGGGACCACTTTCCGGAGGCAAGGC 609
Db 161 SerCysProSerProProGlnArgAlaGlnArgGlyAspHisSerSerTrpGluGlnGly 180
QY 610 CACGCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTCCCGTGGCGGATCTGTG 669
Db 181 GlnAlaProGlyGlySerSerGlnAlaArgGlnIleAspPheProLeuAlaGluValLeuPhe 200
QY 670 CCGACCCAGTTGTTGGTGCATCATCGAAAGAGGCTTGAACATTAAGAACATCACT 729
Db 201 ProThrGlnPheValGlyAlaIleIleGlyLysGluGluLysLeuThrIleLysAsnIleThr 220
QY 730 AAGCAGACCCAGTCCCGGATGATATTCATTAAGAAAGAAACTTGAAGCTGCGAGAGAG 789
Db 221 LysGlnSerArgSerArgValAspIleTyrArgGlnGlnGlnAsnSerArgAlaAlaGluLys 240
QY 790 CTTGTACACATCCATGCCACCCGAGAGGAGCTTCTGAAGCATCCGATGATCTTGAA 849
Db 241 ProValThrMetHisAlaThrProGluGlyThrSerGluAlaCysArgMetIleLeuGlu 260
QY 850 ATCATGAGAAAGAGGAGATGAGACCAACTAGCCGAGAGATCTCTTGAATATCTTG 909
Db 261 IleMetGlnLysGlnAlaAspGluAlaLysLeuAlaGluGluIleProLeuLysIleLeu 280
QY 910 GCACACATGCGCTTGTGTAAGACTGATTGAAAGAGGAGGAGGAGAAATTTGAAGAAATT 969
Db 281 AlaHisAsnGlyLeuValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLysLysAsn 300


```

QY      970 GAACATGAAACAGGAGCCAAAGATACATCTCATCTTTGACAGATTGAGCATATACAAAC 1029
Db      301 GUAHSGlThrGlyThrlYsIleThrlIleSerSerGlnAspSerIleTryasn 320
QY      1030 CCGAAAGAAACATCATCTCTGAAAGGCGCAACAGTTGAGCCTGTGCCAGTCTGAGATAGAG 1089
Db      321 ProGluArgThrIleThrValIlysglyThrValGluValCylAlaSerIleGluIleGlu 340
QY      1090 AATATGAAGAGCTGGCTGAGGCTTGTGAAATGATGATGCTGGCTTTAAACCCACTCC 1149
Db      341 IleMelYsLysLeuArgGluAlaPheGluAsnAspThrLeuThrValAsnThrIleSph 360
QY      1150 GSATATCTTCTCCAGCCTGTACCCCATCACAGTTTGAGCCGCTCCCGCATCATCACTT 1209
Db      361 GlyTyrPheSerSerIleuThrProHlsArgGlnPheGlyProPheProHlsIleSsr 380
QY      1210 TATCCAGAGCAGGAATGATGTGATCTTCTTCAATCCCAACCCAGCTGTGGCGCCATCATC 1269
Db      381 TyrProGluGlnGluIleValAsnLeuPheIleProThrGlnGlyValGlyAlaIleIle 400
QY      1270 GAGAGAAAGGGGCGACATCATCAACAGCTGGCGGATTCGCCGAGCTCTTATCAAGTT 1329
Db      401 GlyTyrSylsglyAlaHlsIleIlysglnLeuAlaArgPheValGlyAlaSerIleIlySle 420
QY      1330 GCCCTGCGGAAAGGCCAGACGTACGCGAAAGATGTATCATCATCACCGGCGCACCGGAA 1389
Db      421 AlaProAlaArgSerPro---LeuArgGlnArgYsValIleIleThrTTrProProGlu 439
QY      1390 GCCCATGTTCAAGCGCCACGAGACGATCTTTGGGAAATCTGAAAGAGAAACTTCTTTAAC 1449
Db      440 SerGlnPheYsAlaGlnGlyArgIlePheGlyYsLeuYsGlnGluAsnPheAsn 459
QY      1450 CCCAAGAGAGAGTGAAGCTGGAAGCGCATATCCAGAGTCCCTTCCACAGCTGGCCGG 1509
Db      460 ProYsGlnAspValYsLeuGlnThrIleHlsIleArgValProSerSerThAlaGlyArg 479
QY      1510 GTGATTGGCAAGGTGGCAGACCCGTGACGAGACTGCGAAGCTTAACAGTGCAGAGATC 1569
Db      480 ValIleGlyYsGlyGlyYsThrValAsnGluLeuGlnAsnLeuIleSerAlaGluVal 499
QY      1570 ATCGTGCCTGCTGACCAAGCCGCAAGATGAATGAGAGATGATGTCAGATTATCGGG 1629
Db      500 IleValProArgAspGlnThrProAspGluAsnGlnGluMetIleValArgIleIleGly 519
QY      1630 CACTTCTTGTGCTACCGAGACTGCACAGCGCAAGATCGGGAATTTGTAACAAACAGTGAAG 1689
Db      520 HisPhePheAlaSerGlnThrAlaGlnArgYsIleArgGluIleValGlnGlnValYs 539
QY      1690 CAGCAGAGAGCAAAATACCTTCAGGAGTGCCTCAGAGCGCAGCAAG 1737
Db      540 GlnGlnGlnGlnGlnYsTyrProGlnGlnGlyValAlaSerGlnArgSerYs 555

```

```

RESULT 6
US-10-313-986-500
; Sequence 500, Application US/10313986
; Publication No. US20030236209A1
; GENERAL INFORMATION:
; APPLICANT: McNabb, Andrea M.
; APPLICANT: Foy, Teresa M.
; APPLICANT: McManabe, Yoshihiro
; APPLICANT: Reed, Steven G.
; APPLICANT: Wang, Tonglong
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C19
; CURRENT APPLICATION NUMBER: US/10/313,986
; NUMBER OF SEQ ID NOS: 560
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 500
; LENGTH: 577
; TYPE: PRF
; ORGANISM: Homo sapiens

```

```

US-10-313-986-500
Alignment Scores:
Pred. No.: 3,34e-148 Length: 577
Score: 1940.00 Matches: 383
Percent Similarity: 79.10% Conservative: 75
Best Local Similarity: 66.15% Mismatches: 95
Query Match: 32.96% Indels: 26
DB: 15 Gaps: 5

US-09-270-437d-8 (1-3283) x US-10-313-986-500 (1-577)
QY      73 ATGAACAAGCTTTATCATGGGAACTGAGACCCCGCTGACCGGACGACCTCCGCGAG 132
Db      1 MetAsnYsLeuYrIleGlyAsnLeuAsnGlnSerValThrProAlaAspLeuYs 20
QY      133 CTCTTGGGGAACGAGAACTGCCCCCTGGCGGAGACGTCCTGCTGAAGTCGGCTTACGCC 192
Db      21 ValPheAlaGlnHlsYsIleSerYsSerGlyGlnPheLeuValYsSerGlyYrAla 40
QY      193 TTGCTGACTACCCCGACAGACTGGGCAATCCGCGCATGAGACCTCTCGGGTAA 252
Db      41 PheValAspCysProAspGlnHlsIleThrAlaMetYsAlaIleGluThrPheSerGlyYs 60
QY      253 GTGCAATTGCAATGGGAAATCATGGAAGTTGATTAATCTGTTAAAGCTTAAGAGAC 312
Db      61 ValGlnLeuGlnGlyYsArgLeuGlnIleGlnHlsSerValProYsYsGlnArgSer 80
QY      313 AGGAAATATTCAGATTTCGAAACATCCCTGCTCAGCTGAGTGGAGGTGTGGATGACTT 372
Db      81 ArgYsIleGlnIleArgAsnIleProProGlnLeuArgTrpGluValLeuAspSerLeu 100
QY      373 TTGCTCATATGAGGACAGTGGAGATGTGGAACAAGTCAACACAGACACAGAAACCGGC 432
Db      101 LeuAlaGlnYrGlyThrValGluAsnCysGlnGlnValAsnThrGlnSerGlnThrAla 120
QY      433 GTTGTCAAGCTCATATGTCAACAAGAGAGAACCAAAATAGCCATGAGAGAGCTAAC 492
Db      121 ValValAsnValThrYrSerAsnArgIleGlnThrArgIleAlaIleMetYsLeuAsn 140
QY      493 GGGCATGATTGGAATCTCTCTTCAAGATTTCTCATCCATCCCGGATGAGAGGTGAGC 552
Db      141 GlnHlsGlnLeuGlnAsnHlsAlaLeuYsValSerYrIleProAspGlnGlnIleAla 160
QY      553 TCCCTTCCGCCCTCCAGGAGCCAGCGTGGGACCACTCTTCCCGGAGACAGGCCAC 612
Db      161 -----GlnGlySerProGluAsnGlnYrArgYsGlyYsIlePheGlySerArgYsGlnProArg 178
QY      613 -----GCCCTGGGGGCACTTCTCAGGCGCAGACAGATTGATTTCCCGTG 657
Db      179 GlnGlySerProValAlaIleAlaGlyAlaProAlaYsGlnGlnGlnValAspIleProLeu 198
QY      658 CGGATCGTGTCCCGACCCGAGTTTGTGTGCCATCATCGGAAAGAGGGCTTACCATTA 717
Db      199 ArgLeuLeuValProThrGlnYrValGlyAlaIleIleGlyYsGlnGlyAlaThrIle 218
QY      718 AAGAAATCACTAAGAGACCCAGCTCCGGGTAGATTCATGAAAGAAAGAACTCTGGA 777
Db      219 ArgAsnIleThrYsGlnThrGlnSerYsIleAspValHlsArgYsGlnAsnAlaGly 238
QY      778 GCTGCAGAGAGGCTGTACCATCATGCGACCCCGAGGGGCACTTCTGAAGACGCGC 837
Db      239 AlaAlaGlnYsAlaIleSerValHlsSerThrProGlnYsSerSerAlaCylYs 258
QY      838 ATGATCTTGAATTCATGCAAGAGGAGAGATGAGAACCAACTAGCCGGAAGATTCCT 897
Db      259 MetIleLeuGlnIleMetHlsYsGlnAlaYsAspThrYsThrAlaAspGluValPro 278
QY      898 CTGAAATCTTGGCACACATGCTGTGGTGAAGCTGATTGGAAGAAAGAGGAGCAAT 957
Db      279 LeuYsIleLeuAlaHlsAsnAsnPheValGlyArgLeuIleGlyYsGlnYsGlnYrAsn 298
QY      958 TTGAAGAAATTCAGATGAAACAGGACCAAGATTAACAATCTTCTTGGAGATTGG 1017

```

[illegible]

Alignment Scores:	1,03e-147	Length:	577
Pred. No.:	1934.00	Matches:	381
Score:	79.10%	Conservative:	77
Percent Similarity:	65.80%	Mismatches:	95
Best Local Similarity:	32.86%	Indels:	26
Query Match:		Gaps:	5
DB:	9		

US-09-270-437D-8 (1-1283) x US-09-873-637-2 (1-577)

QY	ATGAACAAGCTTTAACAATCCGGGAACCTGAGCCCGCGTACCGACCGACGACCTCCGGGAG	132
Db	1 MetasnlyseuYrllleGlyasneAasgluseValThrProAlaspneulnlys	20
QY	133 CTTCTTTGGGAGACGAAGAGTCCCTCGCGGAGACAGCTCTGCTGAAGTCCGGCTAACGCC	192
Db	21 ValhelaaguhstisylsilesertyrsergylglnPheleuVallysersgylYrYala	40
QY	193 TTGCTGACACTACCCGACCAAGAACTGGGCGCATCCGCGCCATGAGAACCCCTCCGGGTAA	252
Db	41 PheValaspcysProAspgrlunhstlrlPameluYsAlaleglunhrPhesersgyllys	60
QY	253 GTGGAATTGCACTGGGAATAATCAAGGAAGTGAATTACTCACTCTCTAAATAAGCTAAGAAC	312
Db	61 ValGlueuglnnglylsArgleuglumeGlnHisSerValProlysylsnglnatgses	80
QY	313 AGGAAATTCAGATTCCAAACATCCCTCCACCTCGACGTGAGGAGTGTGGATGACATT	372
Db	81 ArgylsileglnlleahgsanlleProProglntleuArglrpluValleuAspSerleu	100
QY	373 TTGGCTCAATATGGGACAGTGGAGAAATGTGGAAACAACTCAACACACACACAGAAACCGCC	432
Db	101 leuAlaglnlyrYgltThrValGluAsncYsglnglnAlaasnhrGluSerGlnThrAla	120
QY	433 GTTTCACACGTCCACATATGCAACAAGAGAGAAACAAATAAGCCATGGAGACTTAAGC	492
Db	121 ValValaasnaValThrYrSerasnhrGlnGlnThrArglnAlalleuYlsleuasn	140
QY	493 GGGCATCAGTTTGAAGAACTACTCCTTCAAGATTTCTTCAATCCCGGATGAAGGTGAGC	552
Db	141 GlyhlsnglnleuglnuasnHlsAlaleuYlsValserYrllleProAspGlnGlnlleThr	160
QY	553 TCCCTCTGCCCCCTCAGCGAGCCGACGCTGGGGACACACTCTTCCCGGAGACAGGCGAC	612
Db	161 -----GlnGlyProGlnuasnGlyArgArgGlyYolYpHeGlyserArgGlylnProArg	178
QY	613 -----GCCCTCGGGGGGACCTTCCAGCGCCAGACAGATTGAATTTCCCGCGT	657
Db	179 GlnGlySerProValAlaAlaGlyAlaProAlaYlsnglnGlnProValaAspIleProleu	198
QY	658 CGGATCTGTCGCCACCCAGCTTTGTTGGTGCATCATCGGAAAGAGGCTTGACATA	717
Db	199 ArgleueValaProThrGlnlyrValGlyAlaIlelleGlylsGlnGlyAlaThrIle	218
QY	718 AAGAACATCACTAAGACACCCAGTCCGGGTAAATTCCTAATAAAGAAGAACTCGGA	777
Db	219 ArgAsnllleThrlysglnThrImserYlsleasPValHlsagYlsGlnAsnAlaGly	238
QY	778 GCTCGAGAAAGCCGTGTACCATCATCGACACCCGAGGAGGACTTCTGAAGCATGCGC	837
Db	239 AlaAlaGlnlysaAlaIleSerValHlsSerThrProGlnGlyCybsSerSerAlaCylys	258
QY	838 ATGATTCCTGAATCATGACGAAGAAGGACAGATGAGACCAACTAAGCCGAAGAATTCCT	897
Db	259 MetIleleuglnlleethrIslsYsglnAlaYlsAspThrlystlnAlaAspGlnValPro	278
QY	898 CTGAAAATCTTGGCACACATGGCTTGTTGGAAAGACATGATTGAAAAGAAGGACGAAT	957
Db	279 leuYllylleuAlaHlsAsnAsnPhaValGlyArgleuIleGlylysglnGlyYArgAsn	298
QY	958 TTGAAGAAATTTGAACATGAACAGGGACCAAGATTAACATCTCATCTTTCAGAGATTGG	1017

```

Db      299 LeuLysLysValGluGlnAspThrGluThrLysIleThrIleSerSerLeuGlnAspLeu 318
QY      1018 AGCATATACACCCGGAAGAAACCATCTGTAAGGACGACAGATTGAGCGCTGTGCACAT 1077
Db      319 ThrLeuTyAsnProGlnArgThrIleThrValLysGlyAlaIleGlnSerCysArg 338
QY      1078 GCTGAGATGAGATTATGAAGAACTGCGTGAAGCGCTTTGAAATGATATGCTGCTGTT 1137
Db      339 AlaGluGlnGluIleMetLysLysValArgGluAlaIleGlnAspAlaAlaMet 358
QY      1138 AACACCCACTCC-----GGATACTTC----- 1158
Db      359 SerLeuGlnSerHisLeuIleProGlyLeuAsnLeuAlaAlaGlyLeuPheProAla 378
QY      1159 ---TTCAGCGCTGACCCCATCACAG-----TTTGGCCCGCTTCCCGCAT 1200
Db      379 SerSerSerAlaValProProProProSerSerValThrGlyAlaAlaProTyrSer 398
QY      1201 CATCACTCTTATCCAGACGAGAGATTGTAATCTCTTCATCCCAACCGAGCTGTGGGC 1260
Db      399 PheMetGlnAlaProGluGlnGluMetValGlnValPheIleProAlaGlnAlaValGly 418
QY      1261 GCCATCATCGGGAAGAAAGGGGCGACATCAAAACGCTGGCGAGATTGCGCGAGCCTCT 1320
Db      419 AlaIleIleGlyLysGlyGlyGlnHisIleLysGlnLeuSerArgPheAlaSerIleSer 438
QY      1321 ATCAAGATTGCGCCCGGAGAGCGCCAGACGTCACCGAAAGATGCTCATCATCCCGG 1380
Db      439 IleLysIleAlaProProGluThrProAspSerLysValArgMetValValIleThrGly 458
QY      1381 CCACCGAAGCCAGATTCAAGAGCCGACGAGACGATCTTGGGAAACTGAAAGAGGAAAC 1440
Db      459 ProProGluAlaGlnPheLysAlaGlnIleLysIleLysLysLeuLysGlnLys 478
QY      1441 TTCTTTAACCCTCAAGAAAGAAAGTGAAGTGAAGGCGATTCAGAGTCCCTCTTCCACA 1500
Db      479 PhePheGlyProLysGluGlnValLysLeuGlnThrHisIleArgValProAlaSerAla 498
QY      1501 GCTGGCGGGGTGATTGGCAAGGTGGCAAGCCGTAACGAACTGACAGAACTTAAACAGT 1560
Db      499 AlaGlyArgValIleGlyLysGlyGlyLysThrValLysGlnLeuGlnAsnLeuThrAla 518
QY      1561 GCAGAAATCATCGTGCCTCGTACCAACGCCAGATGCAAAATGAGAAAGTATGATCGTCA 1620
Db      519 AlaGluValValProArgAspGlnThrProAspGlnAsnAspGlnValIleValLys 538
QY      1621 ATTATGGGCACTTCTTGTGTAAGCAAGTGCACAGCGCAAGATCAGGAAATTGTACA 1680
Db      539 IleIleGlyHisPheTyAlaSerGlnMetAlaGlnArgLysIleArgAspIleLeuAla 558
QY      1681 CAGGTGAAGACAGGAGGAGCAAGAAATACCTCAGGAGATCCGCTCAGCGCAGCAAG 1737
Db      559 GlnValLysGlnGlnHisGlnLysGlyGlnSerAsnLeuAlaGlnAlaArgLys 577

```

```

RESULT 8
US-10-313-986-486
; Sequence 486, Application US/10313986
; Publication No. US20030236209A1
; GENERAL INFORMATION:
; APPLICANT: Foy, Teresa M.
; APPLICANT: McNabb, Andria
; APPLICANT: Matanabe, Yoshihiro
; APPLICANT: Reed, Steven G.
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.455C19
; CURRENT APPLICATION NUMBER: US/10/313,986
; NUMBER OF SEQ ID NOS: 560
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 486
; LENGTH: 589

```

```

; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-313-986-486
Alignment Scores:
Pred. No.: 8,14e-144
Score: 1866.00
Percent Similarity: 77.16%
Best Local Similarity: 63.79%
Query Match: 32.04%
DB: 15
Gaps: 8
US-09-270-437D-8 (1-3283) x US-10-313-986-486 (1-589)
QY      73 ATGAACAAGCTTTATCATCGGAGACCTGAGCCCGCGCTACCGCGACGACCTCGGAG 132
Db      1 MetAsnLysLeuTyIleGlyAsnLeuSerGluAsnAlaAlaProSerAspLeuGlnSer 20
QY      133 CTCTTTGGGAGACAGGAAGCTGCGCTGCGGAGACAGGTCTGCTGAAGTCCGCTACGCC 192
Db      21 IlePheLysAspAlaLysIleProValSerGlyProPheLeuValLysThrGlyTyrAla 40
QY      193 TTGCGTGACTACCCCGACGACGAGACTGGGCGATCCGCGCGATCGAGACCTCTCGGATAA 252
Db      41 PheValAspCysProAspGlnSerThrPalaLeuLysAlaIleGluAlaLeuSerGlyLys 60
QY      253 GTGAATTCATGGGAAATCATGAGAGTTGATTACTCAGTCTCTAAAGAGCTAAGAGAGC 312
Db      61 IleGluLeuHisGlyLysProIleGluValGlnHisSerValProLysArgGlnArgIle 80
QY      313 AGGAAATTCAGATTTCGAAACATCCCTCTCTCAGCTGACGTGGAGAGTGTGATGAGACTT 372
Db      81 ArgLysLeuGlnIleArgAsnIleProProHisLeuGlnThrGlyValLeuAspSerLeu 100
QY      373 TTGCGTCAATATGAGGACAGCTGAGAGATGTGAAACAGTCAACAGACACAGACAAACGCC 432
Db      101 LeuValGlnThrGlyValValGlnSerCysGlnGlnValAlaThrAspSerGluThrAla 120
QY      433 GTTGTCACAGCTCACATATGCAACAGAGAAAGCAAAATATGACCATGAGAGCTAAGC 492
Db      121 ValValAsnValThrTyLysSerSerLysAspGlnAlaArgGlnAlaLeuAspLysLeu 140
QY      493 GGGATCATGTTTGAAACTACTCTTCAAGATTTCCTACATCCGCGATGGAAGAGTGAAC 552
Db      141 GlyPheGlnLeuGlnAsnPheThrLeuLysValAlaTyrIleProAspGluThrAlaAla 160
QY      553 TCCCTTGGCGCCCTCGAG-----CGAGCCAGAGT-----GGGACCACTCTCCCGG 600
Db      161 GlnGlnAsnProLeuGlnGlnProArgGlyArgArgGlyLeuGlyGlnArgLysSer 180
QY      601 GAGCAAGCCACAGCCCTCGGAGGACATCTCAGAGCCAGACAGATTGATTTCCGCTGCGG 660
Db      181 ArgGlnGly---SerProGlySerValSerLysGlnLysProCysAspLeuProLeuArg 199
QY      661 ATCCTGATCCCAACCCAGTTGTTGGTGCATCATCGAAAGAGAGGCGCTTGACCATTAAG 720
Db      200 LeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGlnGlyAlaThrIleArg 219
QY      721 AACATCACTAAGAGACCCAGTCCCGGCTGATATCATGAAAGAAAGAACTCTGAGGCT 780
Db      220 AsnIleThrLysGlnThrGlnSerLysIleAspValHisArgLysGlnAsnAlaGlyAla 239
QY      781 GCAAGAGAGCTGTCAACCATCATGACACCCAGAGGAGGACTCTGAGAGCATGCCGATG 840
Db      240 AlaGlnLysSerIleThrIleLeuSerThrProGlnGlyThrSerAlaAlaCysLysSer 259
QY      841 ATTCTTGAATCATGCAAGAAAGGCGAGATGAGACCAAACTAGCCAGAGATTCCTCTG 900
Db      260 IleLeuGlnIleMetHisLysGlnAlaGlnAspIleLysPheThrGlnGluIleProLeu 279
QY      901 AAAATCTGGCAACAATGCGCTGTGGAAGATGATGGAAGAAAGAGCGCAAAATTTG 960
Db      280 LysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGlnGlyArgAsnLeu 299

```

```

QY 961 AAGAAATGAACTGAAACGGGACCAAGATACATCTTGTGAGATTGAGC 1020
Db 300 LysLysIleGluGlnAspThrAspIleThrIleSerProLeuGlnLeuThr 319
QY 1021 ATATACAAACCGGAAAGAACCATCTGTGAAAGGACAGATTGAGCCTGTGCCAGTCT 1080
Db 320 LeuLysAspProGlnArgThrIleThrValLysGlnAsnValGlnThrCysAlaLysAla 339
QY 1081 GAGATGAGATTATGAAAGAGCTGCGTGAAGCCTTTGAAATGATATGCTGCTTAAAC 1140
Db 340 GlnGluGlnIleMetLysIleArgIleSerTYrGlnAsnAspIleAlaSerMetAsn 359
QY 1141 ACCCACTCC-----GATACCTC----- 1158
Db 360 LeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProThr 379
QY 1159 -----TCCAGCTGTACCCCATCAACAGTTTGCCCTTCCGCAT 1200
Db 380 SerGlyMetProProProThrSerGlyProProSerAlaMetThrProProIlyProGln 399
QY 1201 CATCACTCTTATCCAGACGAGAGATTGTGAATCTCTTATCCCAACCGAGCTGTGGC 1260
Db 400 PheGluGln---SerGluThrGluThrValHisLeuPheIleProAlaLeuSerValGly 418
QY 1261 GCCATCATCGGAAAGAGGGGCAACATCAACAAAGCTGCGAGATTGCGCGAGCCTCT 1320
Db 419 AlaIleIleGlyLysGlnGlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSer 438
QY 1321 ATCAAGATTGCCCGCGGAAAGGCGACGTCAGCGAAAGATGCTCATCATCCCGG 1380
Db 439 IleLysIleAlaProAlaGlnAlaProAspAlaLysValArgMetValIleIleThrGly 458
QY 1381 CCACCGGAAACCCAGTTCAAGGCCCAAGGACGAGATCTTTGGAAACTGAAAGAGGAAAC 1440
Db 459 ProProGluAlaGlnPheLysAlaGlnGlyArgIleTYrGlyLysIleLysGlnGluAsn 478
QY 1441 TTCTTTAACCCCAAAAGAAAGTGAAGCTGGAAGGCCATATCAGAGTGCCTTTCCACA 1500
Db 479 PheValSerProLysGlnGluValLysLeuGluAlaHisIleArgValProSerPheAla 498
QY 1501 GCTGCGCGGATGTTGGCAAGAGTGCCAGACCGGTGAACGAACTGCAAACTTAACCACT 1560
Db 499 AlaGlyArgValIleGlyLysGlyLysThrValAsnGluLeuGlnAsnLeuSer 518
QY 1561 GCAGAAATCATGCTGCTCTGTGAACCAACCGCATGAATAAGAAATGATCGTCAGA 1620
Db 519 AlaGluValValAlaProArgAspGlnThrProAspGlnAsnAspGlnValValLys 538
QY 1621 ATATCGGGGACCTTTGTGTAGCGAGACTGCAACAGCGCAAGATCAGGAAATGTACAA 1680
Db 539 IleThrGlyHisPheTYrAlaCysGlnValAlaGlnIleArgLysIleGlnGluIleLeuThr 558
QY 1681 CAGGTGAAGAG---CAGGACGAGAAATACCTTCAG---GGAGTGCCTTCAACGCGCAGC 1734
Db 559 GlnValLysGlnHisGlnGlnGlnIleValAlaLeuGlnSerGlyProProGlnSerArgArg 578
QY 1735 AAGTGAGGCTTCCACAGGCAACGCAAAACAAAC 1767
Db 579 LysHisHisHisHisHisHisHisHisHis 589

```

```

; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Fanger, Neil
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C14
; CURRENT APPLICATION NUMBER: US/09/735,705
; NUMBER OF SEQ ID NOS: 419
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 348
; LENGTH: 579
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-735-705-348

Alignment Scores:
Pred. No.: 6,32e-143 Length: 579
Score: 1875.00 Matches: 369
Percent Similarity: 78.70% Conservative: 78
Best Local Similarity: 64.96% Mismatches: 95
Query Match: 31.86% Indels: 26
Gaps: 6
DB: 9

US-09-270-437d-8 (1-3283) x US-09-735-705-348 (1-579)
QY 73 ATGAAACAAGCTTTACATCGGGAACCTGAGCCCGCGTCACCGCGACGACTCCGCGAG 132
Db 1 MetAsnLysLeuTYrIleGlyAsnLeuSerGluAsnAlaAlaProSerAspLysLeuSer 20
QY 133 CTCTTTGGGAGACGGAAGCTGCCCCCTGCGGAGACAGTCTCTGTAAGTCGGCTACGCG 192
Db 21 IlePheLysAspAlaLysIleProValSerGlyProPheLeuValLysThrGlyTYrAla 40
QY 193 TTCTGTGACTACCCGACACGAACTGGGGCCATCCGCGCATGAGACGCTCTGGGTAA 252
Db 41 PheValAspCysProAspLysLeuIleProValAlaLeuLysAlaIleGluAlaLeuSerGlyLys 60
QY 253 GTGGAATTCATCGGAAATTCATGAAAGTGTACTCACTCACTCTTAAAGCTTAAGAGAGC 312
Db 61 IleGluLeuHisIleGlyLysProIleGluValGlnHisSerValProLysArgGlnArgIle 80
QY 313 AGGAAATTCATGATTCGAAACATCTCTCTACCTGCACTGAGGAGGTGTTGATGACTT 372
Db 81 ArgLysLeuGlnIleArgAsnIleProProHisLeuGlnIleProValLeuAspSerLeu 100
QY 373 TTGGCTCATATGCGGACAGCGGAAATGTGAAACAAGTCAACACAGACAGAAACCGCC 432
Db 101 LeuValGlnTYrGlyValValGlnSerCysGlnGlnValAsnThrAspSerGluThrAla 120
QY 433 GTTGCAACGTCACATATGCAACAGAGAAAGCAAAATATGCCATGTGAGAACTTAAGC 492
Db 121 ValValAsnValThrTYrSerSerLysAspGlnAlaArgGlnAlaLeuAspLysLeuAsn 140
QY 493 GGGGATCATGTTGAAGACTACTCTCTTAAGATTTCCATCACTCCCGGATGAAGAGTGAAGC 552
Db 141 GlyPheGlnLeuGlnAsnPheThrLeuLysValAlaIleTYrIleProAspGlnThrAlaAla 160
QY 553 TCCCTTGGCCCCCTCG-----CGAGCCAGAGT-----GGGACCACTTCCCGG 600
Db 161 GlnGlnAsnProLeuGlnGlnProAlaGlyArgArgGlyLeuGlyGlnArgLysSer 180
QY 601 GAGCAAGCGCACGCGCTGCGGACATCTTCAGGCGACAGACAGATGTGATTCCGCTGCGG 660
Db 181 ArgGlnGly---SerProGlySerValSerLysGlnLysProCysAspLeuProLeuArg 199
QY 661 ATCTGTGTCCTCCACCATGTTGTTGTGTCATCATCGAAAGAGGCTTGAACATAAG 720
Db 200 LeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGlnGlyAlaThrIleArg 219
QY 721 AACATCACTAAGACAGACCCAGCTCCCGGCTAATATCCATAGAAAAGAACTTGGAGCT 780

```

```

RESULT 9
US-09-735-705-348
; Sequence 348, Application US/09735705
; Patent No. US20020052329A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Jiqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hoeken, Nancy
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.

```

Db 220 AsnIlethrIysgInthrInserIysIleAspValHisrIySgluAnaIaGlyAla 239
QY 781 GCAAGAAAGCTGTACCATCCATGCCACCCAGAGGGAGCTTGTGAAGATGCCGATG 840
Db 240 AlAGlulysrIlethrIleuSerThrProgluGlyThrSerIalaIaCylsYser 259
QY 841 ATTCTGAATCTATGCAAGAAAGGAGATGAAACCAATACCCAGAAAGATTCCTCTG 900
Db 260 IleuGluIleuMetIleuMetIleuMetIleuMetIleuMetIleuMetIleuMet 279
QY 901 AAAATCTGSCACAAATGCTGTGGAAGACTGATGGAAGAAAGAGGACAAATTTG 960
Db 280 LysIleuAlaHisAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsn 299
QY 961 AAGAAATTTGACATGAAACAGAGACCAAGATTAACATCTCATCTTTGACGATTTGAC 1020
Db 300 LysIleuIleuGluIleuGluIleuGluIleuGluIleuGluIleuGluIleuGlu 319
QY 1021 ATATACACCCCGGAAAGAAACCATCATCTGTGAAGGACAGTGAAGGCTGTGCCAGTGT 1080
Db 320 LeuTyAsnAsnProgluIleuGluIleuGluIleuGluIleuGluIleuGluIleuGlu 339
QY 1081 GAGATGAGATTTATGAAAGAGCTGCGTGAAGGCTTTGAAATGATATGCTGCTGTTAAC 1140
Db 340 GluGluIleuIleuMetIleuMetIleuMetIleuMetIleuMetIleuMetIleuMet 359
QY 1141 ACCCACTCC-----GGATACCTC----- 1158
Db 360 LeuGluAlaHisIleuIleuProgluIleuAsnIleuAsnAlaIleuGlyLeuPheProThr 379
QY 1159 -----TCGAGCTGTACCCCATCACCAAGTTGGCCGCTTCCGCGAT 1200
Db 380 SerGlyMetProProProProProProSerGlyProProSerAlaMetThrProProGlu 399
QY 1201 CATCATCTTATCCAGAGAGAGATTTGATGATCTCTTATCCCAACCCAGCGCTGTGGC 1260
Db 400 PheGluGlu---SerGluThrGluThrValHisIleuThrIleProAlaIleuSerValGly 418
QY 1261 GCCATCATCGGAGAGAGAGGAGGACACATCAACACGCTGCGAGATTCGCGGAGCTCT 1320
Db 419 AlaIleIleGlySgIngluIleuHisIleuSgInleuSerIlePheAlaGlyAlaSer 438
QY 1321 ATCAAGATTCCTGCTGAGAGGCGCCAGAGCTCAGCCAAAGATGTCATCATACCGGG 1380
Db 439 IleuIleIleAlaProAlaGluAlaProAspAlaLysValAlaMetValIleIleThrGly 458
QY 1381 CCACCGGAGCGGCTTCAGAGGCGCCAGAGCGATCTTTGGGAAAGTGAAGAGAAAC 1440
Db 459 ProProGluAlaGlnPheLysAlaGlnGlyArgIleIleTyIleIleSgInleuAsn 478
QY 1441 TTCTTTAAACCCCAAGAAAGATGAACTGGAACGATATCAGAGTCCCTTCCACA 1500
Db 479 PheValSerProLysSgIngluValLysIleuGluAlaHisIleuValProSerPheAla 498
QY 1501 GCTGCGCGGCTGATTTGCAAAAGTGGCAAGCCGTGAACCACTGCAAGACTTAAACAGT 1560
Db 499 AlaGlyArgValIleGlyLysGlyLysGlyLysThrValAsnGluIleuGluIleuSer 518
QY 1561 GCAGAACTCATCGCTGCGGACCAAGCCAGATGAAATGAGAAAGTGAAGTGTGTGAGA 1620
Db 519 AlaIleuValValValProAlaGspGlnThrProAspGluAsnAsnAsnValValLys 538
QY 1621 ATTATCGGACATCTTTGTGACCAAGCTGCAACAGCGCAAGATCAGGAAATTTGACAA 1680
Db 539 IleThrGlyHisPheTyAlaCysGluValAlaGlnIleuGlySgInleuIleuThr 558
QY 1681 CAGGTGAAGCAGCAGAGCAGAA 1704
Db 559 GlnValLysGlnHisGlnGlnGln 566
RESULT 10
US-09-850-716A-348

; Sequence 348, Application US/09850716A
; Patent No. US20020115139A1
; GENERAL INFORMATION:
; APPLICANT: Kalos, Michael D.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Better, Marc W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C15
; CURRENT APPLICATION NUMBER: US/09/850, 716A
; CURRENT FILING DATE: 2001-05-07
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 348
; LENGTH: 579
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-850-716A-348
Alignment Scores:
Pred. No.: 6,32e-143 Length: 579
Score: 1875.00 Matches: 369
Percent Similarity: 78.70% Conservative: 78
Best Local Similarity: 64.96% Mismatches: 95
Query Match: 31.86% Indels: 26
DB: Gaps: 6
US-09-270-437D-8 (1-3283) x US-09-850-716A-348 (1-579)
QY 73 ATGACAAAGCTTTACATCGGAAAGCTGAGCCCGCGCTCACCGCGAGCACTCCGAGC 132
Db 1 MetAsnLysIleuTyIleGlyAsnIleuSerGluAsnAlaIleProSerAspLeuGluSer 20
QY 133 CTCTTGGGAGCAGAAAGCTGCGCTGGCGGAGAGGTCTCTGTGAAGTCCGCTTACGCC 192
Db 21 IlePheLysAspAlaLysIleProValSerGlyProPheIleuValLysIleGlyTyAla 40
QY 193 TTGCTGACATACCCCGACCAAGACTGGGCGCATCGGCGCATCGAGACCTTCGGGTAAA 252
Db 41 PheValAspLysProAspGluSerThrAlaLeuLysAlaIleGluAlaIleuSerGlyLys 60
QY 253 GTGGAATTGCGAGGAAATCATGCAAGTGAATTACTACCTCTTAAAGCTAAGAGAC 312
Db 61 IleGluIleuHisIleGlySgProIleGluValGluHisSerValProLysArgGlnArgIle 80
QY 313 AGAAATTCAGATTCGAAACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 372
Db 81 ArgLysIleuGlnIleAlaGlnIleProIleProIleIleuGlnIleuGlnIleuAspSerLeu 100
QY 373 TTGCTCAATATGAGGACAGTGGAGATGTGAACTGAACACAGACAGAAACGCC 432
Db 101 LeuValGlnIleuTyAlaValGluSerCysGluIleuIleuAsnThrAspSerGluThrAla 120
QY 433 GTTGTCAAGCTCATATGCAACAGAGAAAGCAAAATAGCCATGAGAGACTAAGC 492
Db 121 ValValAsnValThrTySerSerLysAspGlnAlaArgGlnAlaIleuAspLysLeuAsn 140
QY 493 GGGCATGATTTGAAACTACTCTTCAAGATTTCTTACATCCCGGATGAAGAGTGAAC 552
Db 141 GlyPheIleuGlnIleuGlnAsnPheThrLeuLysValAlaTyIleProAspGluThrAlaAla 160
QY 553 TCCCTTCGCGCCCTCG-----CGAGCCAGCGT-----GGGACCACTTCCCGG 600
Db 161 GlnGlnAsnProIleuGlnIleuProAlaGlyArgArgGlyLeuGlyGlnArgGlySerSer 180
QY 601 GAGCAAGCCAGCCCTGAGGAGCACTTTCAGAGCCAGCAAGATTTATCCCGCTGCGG 660
Db 181 ArgGlnIleu---SerProGlySerValSerLysIleuIleuProCysAspLeuProLeuArg 199
QY 661 ATCTGTCTCCCAACCATGTTGTGTGTCATCATCGAAAGAGAGGCTTGAACATAAG 720
Db 200 LeuIleuValProThrGlnPheValGlyAlaIleIleGlySgIngluAlaThrIleArg 219

```

OY 721 AACATCACTAAGACAGACCAGTCCGCGGTAGATATCATAGAAAAGAGACTCTGGAGT 780
Db 220 AsnIleThrIysGlnThrGlnSerIleAspValHisArgIleGlnHisAlaGly 239
OY 761 GAGAGAAAGCTGTACCATCCATCCACCCAGAGGGAGACTTGTGAAGATGCGCAGT 840
Db 240 AlaGlnIysSerIleThrIleLeuSerThrProGlnGlyThrSerAlaIleCysIysSer 259
OY 841 ATTCTTGAATCATGACAGAAAGGAGAGATGACACCAATGACCGAAGATGCTCTCG 900
Db 260 IleLeuGlnIleMetHisIleLysGlnAlaGlnMetIleLysPheThrGlnGlnIleProLeu 279
OY 901 AAAATCTTGGACACACATGCTGTGGTAGAGACTGATTTGAAAAGAGAGCAAAATTG 960
Db 280 LysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyIleGlnIleArgSerLeu 299
OY 961 AAGAAATTTGAACATGAAACAGGGACCAAGATTAACATCTCATCTTTGGAGATTTGAC 1020
Db 300 LysLysIleGlnGlnAspThrAspThrLysIleThrIleSerProLeuGlnGlnLeuThr 319
OY 1021 ATATACAAACCGGAAAGAACCATCACTGTGAAGGGACACATTTGAGCGCTGTGCTGCT 1080
Db 320 LeuTyrosinProGlnArgThrIleThrValLysGlnAsnValGlnThrCysAlaLysAla 339
OY 1081 GAGATGAGATTATGAAAGAGCTGCGTAGAGCCTTTGAAAATGATATGCTGCTGCTTAC 1140
Db 340 GlnGlnGlnIleMetLysIleArgIleSerTyrglnAsnAspIleAlaSerMetAsn 359
OY 1141 ACCCACTCC-----GGATCTTC----- 1158
Db 360 LeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProThr 379
OY 1159 -----TCCAGCCTGTACCCCGCATCACAGTTTGCCCGTCCCGCAT 1200
Db 380 SerGlyMetProProProThrSerGlyProProSerAlaMetThrProProIyrProGln 399
OY 1201 CATCACTCTTATCCAGACGAGAGATTGTGAATCTTTCATCCCAACCCAGCTGTGGGC 1260
Db 400 PheGlnGln-----SerGlnThrGlnThrValHisLeuPheIleProAlaLeuSerValGly 418
OY 1261 GCATTCATCGGAAAGAAAGGGGACACATCAAAACGCTGGGAGATTGGCCGAGCCTCT 1320
Db 419 AlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSer 438
OY 1321 ATCAAGATTGCCCCCGGAGAGGCGCCAGACGTCAGAAAGATGTGATCATCAACCGG 1380
Db 439 IleLysIleAlaProAlaGlnAlaProAspAlaLysValArgMetValIleIleThrGly 458
OY 1381 CCACCGAAGCCCACTTCAAGGCCCGAGGACGAGATCTTTGGGAACTGAAAGAGAAAC 1440
Db 459 ProProGlnAlaGlnPheLysAlaGlnGlyArgIleTyrglyLysIleLysGlnGln 478
OY 1441 TTCTTTAACCACCAAGAGAGTGAAGTGAAGGCGATATCAAGATGCTCTTCCACA 1500
Db 479 PheValSerProLysGlnGlnValLysLeuGlnAlaHisIleArgValProSerPheAla 498
OY 1501 GCTGGCCGCGTATGGCAAAAGTGGCAAGACCGTGAACGAAGTGCAGAACTTAACAAGT 1560
Db 499 AlaGlyArgValIleGlyLysGlyLysThrValAsnGlnLeuGlnHisLeuSerSer 518
OY 1561 GCAGAGTCATCGTCCCTGTGACCAACCGCCAGATGAATAATGAGAAAGTATCGTCGA 1620
Db 519 AlaGlnValValAlaProArgAspGlnThrProAspGlnAsnAspGlnValValLys 538
OY 1621 ATTATCGGGCACTTCTTGTAGCCAGACTGCACAGCCCAAGATAGAGAAATTGTACAA 1680
Db 539 IleThrGlyHisPheTyraIaCysGlnValAlaGlnArgLysIleGlnGlnIleLeuThr 558
OY 1681 CAGGTGAAGCAGCAGAGCAGAAA 1704
Db 559 GlnValLysGlnHisGlnGlnGln 566

```

```

US-09-897-778-348
; Sequence 348, Application US/09897778
; Patent No. US20020147143A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongfeng
; APPLICANT: Marnierakis, Margarita
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Henderson, Robert A.
; APPLICANT: Peckham, David W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.455C16
; CURRENT APPLICATION NUMBER: US/09/897,778
; CURRENT FILING DATE: 2001-06-28
; NUMBER OF SEQ ID NOS: 467
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 348
; LENGTH: 579
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-897-778-348

Alignment Scores:
Pred. No.: 6,32e-143 Length: 579
Score: 1875.00 Matches: 369
Percent Similarity: 78.70% Conservative: 78
Best Local Similarity: 64.96% Mismatches: 95
Query Match: 31.86% Indels: 26
Gaps: 6

US-09-270-437D-8 (1-3283) x US-09-897-778-348 (1-579)
OY 73 ATGAAACAAGCTTTATCATCGGAGAACCTGAGCCCGCGTCAACGCCGACACTCCGCGAG 132
Db 1 MethanMySleuTyrlleGlyAsnLeuSerGlnAsnAlaIleProSerAspLeuGlnSer 20
OY 133 CTCCTTTGGGAGACAGAGAGCTGCCCTGCGGAGACAGGCTCTGCTGAATCGGCTACGCC 192
Db 21 IlePheLysAspAlaLysIleProValSerIleProPheLeuValLysThrGlyTyraIa 40
OY 193 TTCCGAGACTACCCCGACAGACACTGAGCCGATCCGCGCATCGAGACCTCTCCGGGTAA 252
Db 41 PheValAspCysProAspGlnSerTrpAlaLeuLysAlaIleGlnAlaLeuSerIlyLys 60
OY 253 GTGGAATTGCAATGGGAAATATCATGAAATTGATTACTGATCTCTTAAAGCTAAGAGAC 312
Db 61 IleGlnLeuHisGlyLysProIleGlnValGlnHisSerValProLysArgGlnArgIle 80
OY 313 AGGAAATTTCAAGATTCCAAATATCCCTCTCACTGCTGAGAGCTTTGATGAGACTT 372
Db 81 ArgLysLeuGlnIleArgAsnIleProProHisLeuGlnTrpGlnValLeuAspSerLeu 100
OY 373 TTGGCTCAATATGGGACAGCTGAGAGATGGAACAAGTCAACACAGACAGACAGAAACCGCC 432
Db 101 LeuValGlnTyrglyValValGlnSerCysGlnGlnValAsnThrAspSerGlnThrAla 120
OY 433 GTTGTCACAGTCACATATGACAAACAGAGAAAGCAAAATAGCCATGAGAAAGCTAAGC 492
Db 121 ValValAsnValThrTyrsSerSerLysAspGlnAlaArgGlnAlaLeuAspLysLeuAsn 140
OY 493 GGGCATCAGTTTGAGACTACTCTTCAAGATTTCCTTCAATCCCGGATGAAAGAGTGAGC 552
Db 141 GlyPheGlnLeuGlnAsnPheThrLeuLysValAlaTyrlleProAspGlnThrAlaAla 160
OY 553 TCCCTTGCCCGCTCAG-----CGAGCCAGAGCT-----GGGACCACTCTTCCCGG 600
Db 161 GlnGlnAsnProLeuGlnGlnProArgIleTyraGlyLysGlyGlnArgIleSerSer 180
OY 601 GAGCAAGGCCACGCCCTGTGGGGGCACTTCTCAAGGCCAGACAGATTGATTCGCGCTGGCG 660

```

```
Db      ||||| ..... ||||| ... ||||| |||||
Qy      181 ArgIngly---SerProGlySerValSerIysGlnIysProCysAspLeuProIeuuRg 199
Qy      661 ATCCGATCCCAACCACTTTGTTGTTGCCATCATCGGAAAGAGAGGCTTACCATTAAG 720
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      200 LeuIeuValProthGlnPheValGlyAlaIleIleGlyIysGlnGlyAlaThrIleArg 219
Qy      721 AACATCACTAAGCAGACCCAGTCCCGGTAGATATCCATTAAGAAAGAAACCTCGAGCT 780
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      220 AsnIleThrIysGlnThrGlnSerIleAspValHisArgIysGlnAsnAlaGlyAla 239
Qy      781 GCAAGAAAGCTGTACCATCATCCACCCCAAGAGGAGCTTGAAGCATGCGCATG 840
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      240 AlaGlnIysSerIleThrIleLeuSerThrProGlnGlnThrSerAlaIaCylsYser 259
Qy      841 ATTCTTGAATCAATGAGAAAGAGGAGCATGAGACCAACTAGCCGAAAGATTCCTGTG 900
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      260 IleuGlnIleMetHisIysGlnAlaGlnAspIleIysPheThrGlnIleProIeu 279
Qy      901 AAAATCTGGCACAATGCTGTGGTGAAGAGCTGATTTGAAAAGAAAGGAGAAATTG 960
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      280 LysIleIeuAlaHisAsnAsnPheValGlyArgIleGlyIysGlnGlyArgAsnIeu 299
Qy      961 AAGAAATTTGAACATGAACAGGAGCAGATTAACAATCTCATCTTTGCAAGATTGAGC 1020
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      300 LysIleIleGlnGlnAspThrAspThrIleThrIleSerProIeuGlnIleuThr 319
Qy      1021 ATATACAAACCCGGAAGAAACCATCATCTGTAAGGCGCACAGTTGAGCCTGTGCCAGTGT 1080
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      320 LeuIysAsnProGlnArgThrIleThrValIysGlnValGlnThrCysAlaIysAla 339
Qy      1081 GAGATAGAGATTATGAAGAGAGCTGTGAGGCTTTGAAATGATATGCTGCGCTTAC 1140
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      340 GlnGlnGlnIleIleMetIysIleArgIleuSerIyrglnAsnAspIleIleAsnMetCsn 359
Qy      1141 ACCCATCTC-----GGATCTTC-----GATATCTTC----- 1158
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      360 LeuGlnAlaHisIleuIleProGlyLeuAsnIleuAsnAlaIeuGlyLeuPheProThr 379
Qy      1159 -----TCAGGCTGTACCCCATCAACAGTTGAGCCTGTGCCGAT 1200
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      380 SerGlyMetProProProIleThrSerGlyProProSerAlaMetThrProProIyProGln 399
Qy      1201 CATCATCTTATCATCGAGCAGGAGATTGTGAATCTCTTCCATCCCAACCGAGCTGTGGC 1260
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      400 PheGlnGln---SerGlnThrGlnThrValHisIleuHeIleProAlaIeuSerValGly 418
Qy      1261 GCCATCATCGGAGAAAGAGGAGGCAACATCAACACAGCTGCGAGATTGCGCGAGCTCT 1320
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      419 AlaIleIleGlyIysGlnGlnHisIleIysGlnIleuSerArgPheAlaGlyAlaSer 438
Qy      1321 ATCAAGATTGCGCTGCGGAAAGGCCAGAGCTCAGCGAAAGATGTCTATCATCCGGG 1380
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      439 IleIysIleAlaProAlaGlnAlaProAspAlaIysValArgMetValIleIleThrGly 458
Qy      1381 CCAACCGGAACCCAGATTCAAGGCGCCAGGAGCGATCTTTGGGAAACGGAAGAGAAAC 1440
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      459 ProProGlnAlaGlnPheIysAlaGlnGlyArgIleIyrgIysIleIysGlnGlnAsn 478
Qy      1441 TTCTTAAACCCCAAGAGAGTGAAGCTGGAAGCGCATATCAGAGTCCCTTCCACA 1500
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      479 PheValSerProIysGlnGlnValIysIleuGlnAlaHisIleArgValProSerPheAla 498
Qy      1501 GTTGGCCGGGTGATTGGCAAAAGTGGCAAGACCTGTGAAGCAACTGCGAACTTAACCACT 1560
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      499 AlaGlyArgValIleGlyIysGlyIysThrValAsnGlnIleuGlnIleuSerSer 518
Qy      1561 GCAGAGTCAATCGTGCCTGTGCAAAAGCCAGATGAAGAAATGAGAAAGGATCGTCAGA 1620
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      519 AlaGlnValValValProAspArgPheGlnThrProAspGlnIleuAsnAspGlnValValVal 538
Qy      1621 ATTATGGGGAGCTTTCTTTGCTAAGCAGTCAAGCGCAAGATCAAGGAAATTTGTACA 1680
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db      539 IleThrGlnHisPheTyAlaCysGlnValAlaGlnIleArgIysIleGlnGlnIleuThr 558
Qy      1681 CAGGTGAAGCAGCAGAGCAGAAA 1704
Db      559 GlnValIysGlnHisGlnGlnGln 566

RESULT 12
US-09-897-778-446
: Sequence 446, Application US/09897778
: Patent No. US20020147143A1
: GENERAL INFORMATION:
: APPLICANT: Wang, Tongrong
: APPLICANT: Marnerakis, Margarita
: APPLICANT: Fanger, Gary R.
: APPLICANT: Vedick, Thomas S.
: APPLICANT: Carter, Darick
: APPLICANT: Matanabe, Yoshihiro
: APPLICANT: Henderson, Robert A.
: APPLICANT: Peckham, David W.
: APPLICANT: Fanger, Neil
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: OF TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
: FILE REFERENCE: 210121.455C16
: CURRENT APPLICATION NUMBER: US/09/897, 778
: NUMBER OF SEQ ID NOS: 467
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 446
: LENGTH: 579
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-897-778-446

Alignment Scores:
Pred. No.: 6,32e-143 Length: 579
Score: 1875.00 Matches: 369
Percent Similarity: 78.70% Conservative: 78
Best Local Similarity: 64.96% Mismatches: 95
Query Match: 31.86% Indels: 26
DB: Gaps: 6

US-09-270-437D-8 (1-3283) x US-09-897-778-446 (1-579)

Qy      73 ATGAACAAGCTTTATCATCGGAAACCTGAGCCCGCGCTCACCGCCGACCTCGGAGC 132
Db      1 MetAsnIysLeuIyrrIleGlyAsnIeuSerGlnAsnAlaIaIaProSerAspLeuGlnIuser 20
Qy      133 CTCCTTGGAGCAGGAAGAGCTGCCCTGGCGGAGCAGGTCTCGTGGAAGTCCGGCTAGGCC 192
Db      21 IlePheIysAspAlaIysIleProValSerGlyProPheIeuValIystrGlyTyAla 40
Qy      193 TTGGTGACTACCCCGACACCACTGGGCCATCGCGCCATCGAGACCTCTCGGTTAAA 252
Db      41 PheValAspIySProAspGlnIuserIleProValSerGlyProPheIeuValIystrGlyTyAla 60
Qy      253 GTGGAATTGATGGGAAATATCATGAAAGTTACTAGTCTCTTAAAAGCTAAGAGC 312
Db      61 IleGlnIleuHisGlyIysProIleGlnValGlnHisSerValProIySArgGlnArgIle 80
Qy      313 AGGAAATTTAGATTGGAACATCCCTCCATCGTGAAGGAGGAGTTGGATGACTT 372
Db      81 ArgIysIleuGlnIleThrGAsnIleProIleHisIeuGlnIleIleProIleValIleuAspSerIeu 100
Qy      373 TTGGCTCAATATGAGCAGTGAAGAAATGTGGAACAAGTCAACAGACACAGAAACCGCC 432
Db      101 LeuValGlnIyrcIyValValGlnIuserCysGlnGlnValAsnHisPserGlnThrAla 120
Qy      433 GTTGTCAAGCTCATATGCAACAAGAAAGAAATATGACCATGGAAGCTAAGC 492
Db      121 ValValAsnValThrIySerSerIySAspGlnAlaIeuAspIySleuAsn 140
Qy      493 GGGCATCAGTTTGAAGAACTCTTCAAGATTCTTACATCCCGGATGAAGAGGTAGC 552
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```



```

Db      141 GlyPheGlnLeuGluAsnPheThrLeuIysValAlaIrrilleProAspGluThrAlaIa 160
QY      553 TCCCTTCGCCCTCAG-----CGAGCCAGCGT-----GGGAGCACTCTTCCCGG 600
Db      161 GlnGlnAsnProLeuGlnGlnProArgGlyArgArgGlyLeuGlyGlnArgGlySer 180
QY      601 GAGCAGGCGCAGCGCCCTGGGGGACCTTCTCAGGCCAGACAGATTGTTCCCGCTGGG 660
Db      181 ArgGlnGly---SerProGlySerValSerIysGlnIysProCysAspLeuProLeuArg 199
QY      661 ATCTGTGTCCCAACCCAGTTGTGTGTGGTCATCGAAGAGAGGCGCTTGAACATAAG 720
Db      200 LeuLeuValProThrGlnPheValGlyAlaIleIleGlyIysGlnGlyAlaIrrilleArg 219
QY      721 AACATCATAGCAGACCCAGTCCCGGGTAGATTATCCATGAAAAAGAACTCTTGAGCT 780
Db      220 AsnIleThrIysGlnThrGlnSerIysIleAspValHisArgIysGlnAsnAlaIylala 239
QY      781 GCAGAGAGCGCTGCATCCATCGACCCAGAGGGGACCTTCTGAAGCATGCCGCATG 840
Db      240 AlaGlnIysSerIleThrIleLeuSerThrProGlnGlyThrSerAlaIaIaCysIysSer 259
QY      841 ATTTCTTGAATCATGCAGAAAGAGGACAGATGAGACCAACTAGCCGAGAGATTCTCTG 900
Db      260 IleLeuGlnIleMetHisIleIysGlnAlaGlnAspIleIysPheThrGlnGlnIleProLeu 279
QY      901 AAAATCTTGGCACCAATGGCTTGGTGGAAACATGATTGAAAAAGAGGCGAAATTG 960
Db      280 LysIleLeuAlaHisAsnAsnBheValGlyArgLeuIleGlyLysGlnIylArgAsnLeu 299
QY      961 AAGAAATTTGAACTGAAGAAAGAGGACCAAGATTAACATCTCTTCTTGGAGATTGAG 1020
Db      300 LysIysIleGlnGlnAspThrAspThrIysIleThrIleSerProLeuGlnGlnLeuThr 319
QY      1021 ATATCAACCCCGAAAGAAACCATCACTGTGAAGGGCAGAGCTTGGCCAGTGT 1080
Db      320 LeuTyAsnProGlnArgThrIleThrValIysGlyAsnValGlnThrCysAlaIysAla 339
QY      1081 GAGATGAGATTTATGAAGAGCTGCGTGAGGCTTGAAGAAATGATGCTGGCTGTAC 1140
Db      340 GlnGlnGlnIleMetLysIleArgGlnSerTyGlnAsnAspIleAlaSerMetAsn 359
QY      1141 ACCCACTCC-----GSAATCTTC----- 1158
Db      360 LeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProThr 379
QY      1159 -----TCCAGCCTGTACCCCGCATCACCAAGTTGGCCGCTTCCGCAT 1200
Db      380 SerGlyMetProProProThrSerGlyProProSerAlaMetThrProProTyProGln 399
QY      1201 CATCACTCTTATCCAGAGAGAGATTGGAATCTCTCATCCCAACCCAGGCTGTGGG 1260
Db      400 PheGlnGln---SerIleThrGlnThrValHisLeuPheIleProAlaLeuSerValGly 418
QY      1261 GGCATCATCGGAGAAAGAGGGGACACATCAACAGCTGGCGAGATTTCGCGGAGCTCT 1320
Db      419 AlaIleIleGlyIysGlnGlnIleGlnHisIleIysGlnLeuSerThrPheAlaIaSer 438
QY      1321 ATCAAGATTGCGCTTCGGAAGGCCAGAGCTGACGAAAGAGATGCTCATCACCGGG 1380
Db      439 IleIysIleAlaProAlaGlnAlaProAspAlaIysValArgMetValIleIleThrGly 458
QY      1381 CCACCGGAAGCCAGTTTCAAGGCCCGAGGACCGAGTCTTGGGAACTGAAAGAGAAAC 1440
Db      459 ProProGlnAlaGlnPheLysAlaGlnGlyArgIleTyGlyIysIleLysGlnGlnAsn 478
QY      1441 TTCTTAATCCCAAGAAAGAGAGTGAAGCTGAAGCGCATATCAGATGCCCTTCCACA 1500
Db      479 PheValSerProLysGlnGlnValIysLeuGlnAlaHisIleArgValProSerPheAla 498
QY      1501 GCTGGCCGGGTGATTGGCAAAGGTGGCAGACCGTGTAACGAACTGCAGAACTTAACAGT 1560
Db      499 AlaGlyArgValIleGlyLysGlyIysThrValAsnGlnLeuGlnAsnLeuSerSer 518

```

```

QY      1561 GCAGAGATCGATGCTGCTGTGACCAAGCCAGATGAATAATGAGGAAGTATGCTCAGA 1620
Db      519 AlaGlnValValProArgAspGlnThrProAspGlnAsnAspGlnValValIys 538
QY      1621 ATTAATCGGGACATTCTTGTGTAAGCCAGACTGCACAGCCCAAGATCAGGAAATGTACA 1680
Db      539 IleThrGlnHisPheTyAlaCysGlnValAlaGlnArgIysIleGlnGlnIleLeuThr 558
QY      1681 CAGGTGAAGCAGCAGGACAGAAA 1704
Db      559 GlnValIysGlnHisGlnGlnGln 566

RESULT 13
US-09-897-778-449
; Sequence 449: Application US/09897778
; Patient No. US20020147143A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Marnierakis, Margarita
; APPLICANT: Fanger, Gary R.
; APPLICANT: Veddyck, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Henderson, Robert A.
; APPLICANT: Peckham, David W.
; APPLICANT: Fanger, Neil
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C16
; CURRENT APPLICATION NUMBER: US/09/897,778
; CURRENT FILING DATE: 2001-06-28
; NUMBER OF SEQ ID NOS: 467
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 449
; LENGTH: 579
; TYPE: PRN
; ORGANISM: Homo sapiens
US-09-897-778-449

Alignment Scores:
Pred. No.: 6,32e-143 Length: 579
Score: 1875.00 Matches: 369
Percent Similarity: 78.70% Conservative: 78
Best Local Similarity: 64.96% Mismatches: 95
Query Match: 31.86% Indels: 26
DB: 9 Gaps: 6

US-09-270-437D-8 (1-3283) x US-09-897-778-449 (1-579)
QY      73 ATGAACAAGCTTTACATCGGAACTGAGCCCGCGCTACCGCGCAGACTCCGAGAG 132
Db      1 MetAsnLysLeuTyIleGlyAsnLeuSerGlnAsnAlaAlaProSerAspLeuGlnSer 20
QY      133 CTCCTTGGGACAGAGAAGCTGCCCTGGCGGACAGAGTCTGTGTAAGTCCGCTAGGCC 192
Db      21 IlePheLysAspAlaLysIleProValSerGlyProPheLeuValIysThrGlyTyAla 40
QY      193 TTCTGGAGTACCCCGACCAAGAACTGGGCGCATCGGCGCATCGAGACCCCTCTGGGTAA 252
Db      41 PheValAspCysProAspIleSerThrAlaLeuLysAlaIleGlnAlaLeuSerGlyLys 60
QY      253 GTGGAATTCATCGGAGAAATCATGGAAGTGTACTAGTCTCTAATAAGCTTAAGAGG 312
Db      61 IleGlnLeuHisGlyLysProIleGlnValGlnHisSerValProLysArgGlnArgIle 80
QY      313 AGGAATAATTCAGATTGGAACATCCCTCTCACTGCAATGGAGGAGGTGTTGATGAGCTT 372
Db      81 ArgIysLeuGlnIleArgAsnIleProProHisLeuGlnIleProLysValIleuAspSerLeu 100
QY      373 TTGGTTCATATGAGGACAGTGGAGATGTGGAACAAGTCAACAGTCAACAGACAGAAACGCC 432
Db      101 LeuValGlnTyGlyValValGlnIysSerCysGlnGlnValAsnThrAspSerGlnThrAla 120

```

```
QY 433 GTTTCACGTCACATATGCAACAAAGAGAGAGCAAAAATAGCCATGAGAGACTTAAC 492
Db 121 ValValAsnValThrTyrSerSerTyrAspGlnAlaArgGlnAlaLeuAspTyrLeuAsn 140
QY 493 GGGGCACTAGTTGAGAACTACTCTTCAAGATTCTCCATCATCCCGATGAAGAGGTGAGC 552
Db 141 GlyPheGlnLeuGlnAsnPheThrLeuTyrValAlaTyrTyrLeuProAspLysThrAlaAla 160
QY 553 TCCCTTCGCCCCCTCAG-----CGAGCCAGCGCT-----GGGACCACTCTTCCCGG 600
Db 161 GlnGlnAsnProLeuGlnGlnProArgGlnTyrArgGlnTyrLeuGlnTyrAlaArgTyrSer 180
QY 601 GAGCAAGGCCAGCCCTGCGGAGCACTTCTCAGGCCAGACAGATTGATTCCCGCTGGGG 660
Db 181 ArgGlnTyr---SerProGlySerValSerTyrGlnTyrProCysAspLeuProLeuArg 199
QY 661 ATCTCGGTTCCTCCACCCAGCTTTGTTGTGTCATCATCGGAAAGAGAGGCTTGAACATAAG 720
Db 200 LeuLeuValProThrGlnPheValGlnAlaTyrLeuGlnTyrGlnTyrAlaThrThrLeuArg 219
QY 721 AACATCACTAAGCAGACCCAGCCCGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT 780
Db 220 AsnThrTyrGlnThrGlnThrGlnThrTyrGlnThrTyrGlnThrTyrGlnThrTyrGlnThr 239
QY 781 GCAGAGAGAGCCCTGACCATCATGCAACCCAGAGGAGGAGCTTGAAGCATGCGCGCATG 840
Db 240 AlaGlnTyrSerTyrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 259
QY 841 ATCTTGAATCATGCGAGAAAGAGGAGAGATGAGCAAACTAGCCGGAAGAGATTCTCTG 900
Db 260 ThrLeuGlnThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 279
QY 901 AAAATCTTGGCAGCAATGCTTGTGTTGGAAGATGATGTTGAAAAGAGAGGAGAAATTTG 960
Db 280 LysThrLeuAlaHisAsnAsnPheValGlnTyrLeuThrGlnTyrGlnTyrGlnTyrGlnTyr 299
QY 961 AAGAAATTTGAATGAGAAACAGAGAGAGATTAACAATCTGATCTTTGACAGATTGAGC 1020
Db 300 LysTyrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 319
QY 1021 ATATACAAACCCGGAAGAGACATATGTTGAAAGGAGAGCTTGAAGGCTGTGCGCATGCT 1080
Db 320 LeuTyrAsnProGlnTyrGlnThrThrThrThrThrThrThrThrThrThrThrThrThrThr 339
QY 1081 GAGATAGATTTATGAGAGAGAGCTGCGAGAGCTTTGAAAATGATATGCTGCTGTATAC 1140
Db 340 GlnThrGlnThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 359
QY 1141 ACCCACTGC-----GGATATCTTC----- 1158
Db 360 LeuGlnAlaHisLeuThrLeuProGlyLeuAsnLeuAsnAlaLeuGlnTyrLeuPheProThr 379
QY 1159 -----TCGAGCTGTACCCCATCACCAAGTTTGCCGTTCCCGCAT 1200
Db 380 SerGlyMetProProProProThrSerGlyProProSerAlaMetThrProProTyrProGln 399
QY 1201 CATCATCTTATTCAGAGAGAGAGATTTGATATCTTCAATCCCAACCCAGGCTGTGGC 1260
Db 400 PheGlnGln---SerGlnThrGlnThrValHisLeuPheThrProAlaLeuSerValGln 418
QY 1261 GCCATCATCGGAGAGAGAGGAGGAGACATCAAAACAGCTGCGAGATTGCGCGAGCTCT 1320
Db 419 AlaThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 438
QY 1321 ATCAAGATTGCTCCCTGCGAGAGGCCCAAGAGTCAAGGAGAAAGTGTATCATCATCCGGG 1380
Db 439 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 458
QY 1381 CCACCGGAGGCCAGTTCAAGGCCAGGAGCGATCTTTGGGAAAGCTGAAAGAGGAGAAAC 1440
Db 459 ProProGlnAlaGlnPheTyrAlaGlnTyrArgTyrThrTyrThrTyrThrTyrThrTyrThr 478
```

```
QY 1441 TTCTTAAACCCCAAGAGAGAGTGAAGCTGAAGGCGATATCAAGAGTGCCTCTTCCAG 1500
Db 479 PheValSerProTyrGlnGlnValTyrLeuGlnAlaHisThrTyrValProSerPheAla 498
QY 1501 GCTGCGCGGATGATTTGAGAAAGTGTGCAAGACCGGAGAGCAAACTGACAGAACTTAACCG 1560
Db 499 AlaGlnTyrGlnValThrGlnTyrGlnTyrGlnTyrGlnTyrGlnTyrGlnTyrGlnTyrGln 518
QY 1561 GCAGAAATGATCTGTCCTCGTGAACCAACGCCAGATGAAGAAATGAGAAAGTGAATGCTCAG 1620
Db 519 AlaGlnValValProArgAspThrThrProAspGlnAsnAspGlnValValTyrVal 538
QY 1621 ATTATCCGGCACTTTGTTGCTAGCCAGACTGACAGCCGAGATGAGGAAATTTGATCAA 1680
Db 539 ThrThrGlnThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 558
QY 1681 CAGGTGAAGCAGAGAGAGAGAGAA 1704
Db 559 GlnValTyrGlnThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 566

RESULT 14
US-10-007-700-348
; Sequence 348, Application US/10007700
; Publication No. US20030064947A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Li, Samuel X.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Fanger, Neil
; APPLICANT: Retter, Marc W.
; APPLICANT: Durham, Margareta
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Peckman, David W.
; APPLICANT: Cai, Feng
; APPLICANT: Foy, Teresa M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C17
; CURRENT APPLICATION NUMBER: US/10/007,700
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 348
; LENGTH: 579
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-007-700-348

Alignment Scores:
Pred. No.: 6.32e-143 Length: 579
Score: 1875.00 Matches: 369
Percent Similarity: 78.70% Conservative: 78
Best Local Similarity: 64.96% Mismatches: 95
Query Match: 31.86% Indels: 26
DB: 12 Gaps: 6

US-09-270-437D-8 (1-3283) x US-10-007-700-348 (1-579)
QY 73 ATGAGCAAGCTTACATCGGAGAGCTGAGCCCGCGTACCGCGAGCACTTCGGGAG 132
Db 1 MetAsnTyrLeuTyrThrGlnAsnLeuSerGlnAsnAlaProSerAspLeuGlnSer 20
QY 133 CTCCTTGGGAGACAGAGAGCTGCGCTGCGGAGAGAGAGTCTGTCGTAAGTCCGGCTACGCC 192
Db 21 ThrPheTyrAspAlaTyrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 40
```

```

QY 193 TTGGTGAACACCCGACAGAACTGGGCGCATCCGGCCATCGAAGCCCTCTGGGGTAA 252
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41 PheValaspCysProaspGlnSerTyrPAlaLeuysAlaIleGlnAlaLeuSerGlyLys 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 253 GTGGAATTCATGGGAAAATATGAAAGTTGATTACTAGTCTCTAAAAAGCTAAGAGC 312
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 IleGlnLeuHISGlyLysProIleGlnValGlnHISSerValProLysArgGlnArgIle 80
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 313 AGGAAATTCAGATTGGAACATCCCTCTCTCACTGCGAGTGGAGGTGTGGATGACTT 372
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 81 ArgLysLeuGlnIleArgAsnIleProProHISLeuGlnTyrGlnValLeuAspSerLeu 100
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 373 TTGGCTCAATTCGGACACGTGGAATGTGAACAAGTCAACACAGACAGAAACCGCC 432
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 101 LeuValGlnTyrGlyValValGlnSerCysGlnGlnValAsnThrAspSerGlnThrAla 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 433 GTTGTCACGTCACATATGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 492
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 ValValAsnValThrTyrSerSerLysAspGlnAlaArgGlnAlaLeuAspLysLeuAsn 140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 493 GGGCATCAGTTTGAGAACTACTCTTCAAGATTCTCTACATCCCGATGGAAGAGTGAGC 552
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 141 GlyPheGlnLeuGlnAsnPheThrLeuLysValAlaTyrIleProAspGlnThrAlaAla 160
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 553 TCCCTTGGCCCCCTCAG-----CGAGCCAGCGT-----GGGAGACCATCTTCCCGG 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 161 GlnGlnAsnProLeuGlnPheProArgGlyArgGlyLeuGlyGlnArgGlySerSer 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 601 GAGCAAGGCGACGCGCTGGGGGCACTTCTCAGGCGCAGACAGATTGATTTCCCGTGGC 660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 ArgGlnGly---SerProLysSerValSerLysGlnLysProCysAspLeuArg 199
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 661 ATCTGTGTCACCAACCAAGTTTGTGTGTCATCATCGAAGAGAGAGGCTTGACCAATAG 720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 200 LeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGlnGlyAlaThrIleArg 219
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 721 AACATCACTAAGCAACCCAGTCCGGGAGATATCCATGAGAAAGAACTCTGAGACT 780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 220 AsnIleThrLysGlnThrLysSerLysIleAspValHisArgLysGlnAsnAlaGlyAla 239
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 781 GCAGAGAGACCTGTCCACATCCATGCCACCCGAGAGGAGACTTCTGAGCATGCCGATG 840
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 240 AlaGlnLysSerIleThrIleLeuSerThrProGlnGlyThrSerAlaIaCysLysSer 259
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 841 ATCTTGAAATCATGCAAGAAAGGCAAGATGAGACCAAACTAGCCGAAAGATCTCTCG 900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 260 IleLeuGlnIleMetHISGlyAlaGlnAspIleLysPheThrGlnGlnIleProLeu 279
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 901 AAAATCTTGCAACATGGCTGTGTGGAAGCTAGTTGGAAGAAAGAGCAAAATTTG 960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 280 LysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGlnGlyArgAsnLeu 299
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 961 AAGAAATTCAGATGAAACAGAGGAGCAAGATTAACATCTCATCTTTGACGATTTGAGC 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 300 LysLysIleLeuGlnAsnThrAspThrLysIleThrIleSerProLeuGlnGlnLeuThr 319
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1021 ATATACACCCGAAAGAACATCATCTGTGAAGGCACTTGAAGCGCTTGCGAGTGC 1080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 320 LeuTyrAsnProGlnTyrGlnThrIleThrValLysGlyAsnValGlnThrCysAlaLysAla 339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1081 GAGATGAGATTATGAAAGAGCTGCGGAGGCTTTTAAATGATATGCTGCGCTGTAAAC 1140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 340 GlnGlnIleLeuMetLysLysIleArgLysSerTyrGlnAsnAspIleAsnMetAsn 359
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1141 ACCCACTCC-----GGATACCTTC----- 1158
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 360 LeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProThr 379
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1159 -----TCCAGCTGTACCCCATACACAGATTGGCCGCTCCGCGCAT 1200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 380 SerGlyMetProProThrThrSerGlyProProSerAlaMetThrProProTyrProGln 399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1201 CATCACTCTTATCCAGAGCAGAGATTTGTGAATCTTTCAATCCCAACCCAGGCTGTGGGC 1260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 400 PheGlnGln---SerGlnThrGlnThrValHisLeuHISLeuProAlaLeuSerValGly 418
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1261 GCCATCATCCGGAGAGAGAGGCGCACATCAACCAAGCTGGCGAGATTCCGCGAGCTCT 1320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 419 AlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSer 438
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1321 ATCAAGATTGCCCTCGCGAAGGCCACAGCTCAACGAAAGATGCTCATATCAACCGG 1380
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 439 IleLysIleAlaProAlaGlnAlaProAspAlaLysValArgMetValIleIleThrGly 458
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1381 CCACCGGAGGCCCACTTAAAGGCCCGAGAGAGATCTTTGGGAAATCGAAAGAGAAAC 1440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 459 ProProGlnAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLysGlnAsn 478
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1441 TTCTTTAACCACCAAGAGAGAGTGAAGCTGGAAGCGCATATCAGAGTCCCTTCCACA 1500
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 479 PheValSerProLysGlnGlnValLysLeuGlnAlaHisIleArgValProSerPheAla 498
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1501 GCTGCGCGGTGATTGGCAAAAGGTGGCAAGACCGTGAACGAATCGACAACTTAAACAGT 1560
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 499 AlaGlyArgValIleGlyLysGlyGlyLysThrValAsnGlnLeuGlnAsnLeuSerSer 518
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1561 GCAGAAATCATGCTGCTCTCGTGAACCAAGCCAGATGAAATGAGAGAGTATCGTCA 1620
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 519 AlaGlnValAlaValProArgAspGlnThrProAspGlnAsnAspGlnValValLys 538
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1621 ATTATCGGGCACTTTGCTTGGCCAGACTGCAACGCGCATATCAGAGATGAGGAAATTGTACA 1680
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 539 IleThrGlyHisPheThrAlaCysGlnValAlaGlnIleArgLysIleGlnGlnIleLeuThr 558
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1681 CAGGTGAAGCAGCAGAGAGCAAGAA 1704
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 559 GlnValLysGlnHisGlnGlnGln 566
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 15
US-10-007-700-446
; Sequence 446, Application US/10007700
; Publication No. US20030064947A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Wang, Aijun
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Li, Samuel X.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Ranger, Neil
; APPLICANT: Retter, Marc W.
; APPLICANT: Durham, Margarita
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Derrick
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Beckman, David W.
; APPLICANT: Cai, Feng
; APPLICANT: Fey, Teresa M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C17
; CURRENT APPLICATION NUMBER: US/10/007,700
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 446
; LENGTH: 579
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-007-700-446

```

```

Alignment Scores: 632e-143 Length: 579
Pred. No.: 1875.00 Matches: 369
Score:

```

Percent Similarity: 78.704
 Best Local Similarity: 64.964
 Query Match: 31.864
 DB: 12
 Gaps: 6
 US-09-270-437D-8 (1-3283) x US-10-007-700-446 (1-579)

```

QY 73 ATGAAACACCTTTACATGGGAAACCTGAGCCCGCGGTACCGCGACGACTCCGGCAG 132
DB 1 MetAsnLysLeuTrpIleGlyAsnLeuSerGluAsnAlaIleProSerLeuGluSer 20
QY 133 CTCCTTTGGGAGACGAGAGCTGCGCCCTGCGGAGACAGTCTCTGTAAGTCCGCTACGCC 192
DB 21 IlePheLysAspAlaLysIleProValSerGlyProPheLeuValLysThrGlyTyrAla 40
QY 193 TTCTGGGACTACCCCGACACCAACTGGCGCATCCGCGCATGAGAACCTCTCGGCTAAA 252
DB 41 PheValAspCysProAspGluSerThrPalaLeuLysAlaIleGluAlaLeuSerGlyLys 60
QY 253 GTGGAATTGACATGGGAAATCATGGAAGTGAATTACTAGTCTTAAAAAGCTAAGAGAC 312
DB 61 IleGluLeuHisGlyLysProIleGluValGluHisSerValProLysArgGluArgIle 80
QY 313 AGGAAATATTCAGATTGAAACATCCCTCTCCTCAGCTGAGGAGAGTGTGATGAGACTT 372
DB 81 ArgLysLeuGlnIleArgAsnIleProProHisLeuGlnIleProGluValLeuAspSerLeu 100
QY 373 TTGGCTCAATATGGGACAGTGGAGAAATGTGAAACAGTCAACACAGACAGAAACCGCC 432
DB 101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSerGluThrAla 120
QY 433 GTTGTCACGTCACATATGCAACAGAGAAAGCAAAATAGCATGAGAGAGTAAAGC 492
DB 121 ValValAsnValThrTyrSerSerLysAspGlnAlaArgGlnAlaLeuAspLysLeuAsn 140
QY 493 GGGCATCAGTTTGAAGACTACTCTTCAAGATTCTCTCATCCCGAGTGAAGAGGTGAGC 552
DB 141 GlyPheGlnLeuGluAsnPheThrLeuLysValAlaTyrIleProAspGlnThrAlaAla 160
QY 553 TCCCTCTGGCCCTCAG-----CGAGCCCAAGCGT-----GGGACCACTCTTCCCGG 600
DB 161 GlnGlnAsnProLeuGlnGlnProArgGlyArgArgGlyLeuGlnArgGlySerSer 180
QY 601 GAGCAAGGCCACGCGCCCTGGGGGCACTTTCAGCGCCAGACAGATGATTTCCCGCTGCG 660
DB 181 ArgGlnGly--SerProGlySerValSerLysGlnLysProCysAspLeuProLeuArg 199
QY 661 ATCTGTGTCCTCCACCCAGTTTGTGTGTCATCATCCGAAAGAGAGGCTTGACCATTAAG 720
DB 200 LeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAlaThrIleArg 219
QY 721 AACATCACTAAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAACTCTGAGCT 780
DB 220 AsnIleThrLysGlnThrGlnSerLysIleAspValHisArgLysGluAsnAlaGlyAla 239
QY 781 GCAGAGAAGCCTGCACATCCATGCAACCCCAAGAGGAGATTCTGAAGCATGCCCATG 840
DB 240 AlaGluLysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCysLysSer 259
QY 841 ATTCTTGAATTCATGACAGAAAGAGAGCATAGACCAACTAGCCGAAGAGATTCCCTGTG 900
DB 260 IleLeuGluIleMetHisLysGluAlaGlnAspIleLysPheThrGluGluIleProLeu 279
QY 901 AAAATCTTGGCACAATGGCTTGGTGAAGACTGATTGAAAAAGAGCGAATTTG 960
DB 280 LysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeu 299
QY 961 AAGAAATTTGAACATGAAACAGGAGCAAGATTAACATTCATCTTGGCAGATTTGAGC 1020
DB 300 LysLysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnGluLeuThr 319
QY 1021 ATATACAAACCCGAAAGAACCATCATGTGAAGGGCACAGTTGAGGCTGTGACGTGCT 1080
DB 1081 GAGATAGAGATTATGAAAGAGCTGCGTGAAGCCCTTTGAAATGATGCTGCTGTTAAC 1140
DB 340 GlnGluGluIleMetLysLysIleArgGluSerTyrGluAsnAspIleAlaSerMetAsn 359
QY 1141 ACCCACTCC-----GGAATCTTC----- 1158
DB 360 LeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProProThr 379
QY 1159 -----TCCAGCCTGTACCCCTTACACAGATTGGCCCTTCCCGCAT 1200
DB 380 SerGlyMetProProProThrSerGlyProProSerAlaMetThrProProTyrProGln 399
QY 1201 CATCACTCTTATCCAGACGAGAGATTGGAATCTCTTCATCCCAACCAAGCGTGGCGGC 1260
DB 400 PheGluGln---SerGlnThrGluThrValHisPhePheIleProAlaLeuSerValGly 418
QY 1261 GCCATCATCGGAGAGAGAGGCGGACACATCAAAACAGCTGGCGAGATTGGCGGACCTCT 1320
DB 419 AlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSer 438
QY 1321 ATCAAGATTGGCCCTCGGAGAGCCCAAGACGTCAAGCGAAAGATGTGATCATCAACCGG 1380
DB 439 IleLysIleAlaProAlaGluAlaProAspAlaLysValArgMetValIleIleThrGly 458
QY 1381 CCACCGAAGCCAGTTCAAGGCGCAGAGCGATCTTGGGAAACTGAAAGAGAAAC 1440
DB 459 ProProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLysGluGluAsn 478
QY 1441 TTCTTTAAACCCCAAGAAAGAGTGAAGCTGGAAGCGCATATCAGAGTCCCTTCCACA 1500
DB 479 PheValSerProLysGluGluValLysLeuGlnAlaHisIleArgValProSerPheAla 498
QY 1501 GCTGCGCGGCGATTTGGCAAGGTGGCAAGCCGTGAACCAACTGACCAATCTTAACAGT 1560
DB 499 AlaGlyArgValIleGlyLysGlyLysThrValAsnGluLeuGlnAsnLeuSerSer 518
QY 1561 GCAGAGTCATCGTGCCTCGTGAACCAACGCGCATGAAATGAGGAAGTATGATGTCAG 1620
DB 519 AlaGluValValValProArgAspGlnThrProAspGluAsnAspGlnValValLys 538
QY 1621 ATTATCGGCGCATCTTCTTGTAGCCCAAGCTGACAGCGCAAGATCAGGAAATGTACAA 1680
DB 539 IleThrGlyHisPheTyrAlaCysGlnValAlaGlnArgLysIleGlnIleLeuThr 558
QY 1681 CAGGTGAAGACGACGAGCAAGAAA 1704
DB 559 GlnValLysGlnHisGlnGln 566

```

Search completed: July 23, 2004, 11:58:03
 Job time : 192.647 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_nzp model

Run on: July 23, 2004, 10:55:36 ; Search time 49.5997 Seconds
(without alignments)
12733.810 Million cell updates/sec

Title: US-09-270-437D-8
Perfect score: 5886
Sequence: 1 ggcagcgagagcgagcgagga.....aaccttgtaaatgttattt 3283

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+nzp.model -DBY=xlp
-O=/cgg2/1/USPRO_spool/p/US09270437/runat_23072004_095740_27439/app_query.fasta_1.5582
-DB=PIR_78 -QWRT=fastan -SUFFIX=xpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=prco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USR=US09270437 @CGN 1.1.134 @runat_23072004_095740_27439 -NCP=6 -ICPU=3
-NO_MMAP -IARROTARY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -MAIN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	474.5	8.1	854	2	T23837	hypothetical prote
2	274	4.7	589	2	T19216	hypothetical prote
3	274	4.7	611	2	T19217	hypothetical prote
4	272	4.6	568	2	T49962	hypothetical prote
5	263.5	4.5	621	2	D96554	hypothetical prote
6	262.5	4.5	644	2	A53184	hypothetical prote
7	261.5	4.4	768	2	T27855	myc far upstream e
8	254	4.3	680	2	T25832	hypothetical prote
9	250	4.2	398	2	T41600	hypothetical prote
10	240	4.1	313	2	T48439	probable pre-mRNA
11	234.5	4.0	479	2	C86275	hypothetical prote
12	229.5	3.9	413	2	S46109	hnRNP complex prot
13	225	3.8	356	2	S58529	hypothetical prote
14	224	3.8	846	2	T04533	hypothetical prote

15	223	3.8	1268	2	A44125	high density lipop
16	211.5	3.6	362	2	S78515	single-stranded nu
17	210.5	3.6	632	2	T02627	hypothetical prote
18	210	3.6	365	2	S42471	hypothetical prote
19	209.5	3.6	397	2	T30168	hnRNP protein E2 -
20	208.5	3.5	396	2	S41224	hypothetical prote
21	207	3.5	510	2	T18489	hnRNP protein - Af
22	207	3.5	1270	2	S23464	oncogene ventral
23	195.5	3.3	1279	2	T41389	vigilin - chicken
24	195	3.3	1270	2	T19117	rna binding protei
25	194	3.3	1198	2	B88279	hypothetical prote
26	186	3.2	748	2	T04011	protein COB89.2 (i
27	184	3.1	629	2	T05425	hypothetical prote
28	183.5	3.1	1283	2	T19652	polyadenylate-bind
29	180	3.1	806	2	T13690	related to SCP160
30	180	3.1	1222	2	S56030	hypothetical prote
31	174	3.0	464	2	S43363	SCP160 protein - Y
32	174	3.0	464	2	A54143	transformation upr
33	173.5	2.9	924	2	S27923	kappa-B motif-bind
34	173	2.9	463	2	S41495	gene Lf3 protein -
35	172	2.9	660	1	Q0BE3	dc stretch-binding
36	170.5	2.9	381	2	S45766	BHLF1 protein - hu
37	169.5	2.9	968	2	T00353	hypothetical prote
38	169.5	2.9	1684	2	JW0057	hypothetical prote
39	169	2.9	495	2	T04255	gravin - human
40	162.5	2.8	473	2	A54691	hypothetical prote
41	162	2.8	1791	2	T02345	octamer-binding pr
42	161.5	2.7	471	2	S41768	hypothetical prote
43	161.5	2.7	1003	2	C84601	splicing factor ho
44	161	2.7	3938	2	T42761	hypothetical prote
45	160.5	2.7	471	2	G01211	Bassoon protein - h
						54 kDa protein - h

ALIGNMENTS

RESULT 1
T23837
hypothetical protein M88.5 - Caenorhabditis elegans
C:/Species: Caenorhabditis elegans
C:/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:/Accession: T23837
R:/Sulston, J.
submitted to the EMBL Data Library, June 1994
A:/Reference number: Z19806
A:/Accession: T23837
A:/Status: preliminary; translated from GB/EMBL/DBJ
A:/Molecule type: DNA
A:/Residues: 1-854 <MIL>
A:/Cross-references: EMBL:Z34802; PTDN:CAA84338.1; GSPDB:GN00021; CESP:M88.5
A:/Experimental source: clone M88
C:/Genetics:
A:/Gene: CESP:M88.5
A:/Map position: 3
A:/Introns: 38/3; 96/1; 133/3; 255/3; 307/3; 395/2; 436/2; 492/3; 657/3; 759/1; 820/1

Alignment Scores:

Pred. No.: 1.48e-24
Score: 474.50
Length: 854
Percent Similarity: 40.17%
Matches: 153
Best Local Similarity: 25.71%
Conservative: 86
Query Match: 8.06%
Mismatches: 176
Indels: 181
Gaps: 18

US-09-270-437D-8 (1-3283) x T23837 (1-854)

QY	422	CAGAAACCGCGCTGTGTCACATGATGCAACAGA-----	459
DB	244	GLNHlProlgImetvctglnghlnhSAlaGlnGlnGlyTyrhIspROIhGlnGlnAsn	263
QY	460	GAAAGAGCAAAATAACCATGAGAAAGCTAAGCGGCATCATGTTGAGAACTACCTCCTTC	519
DB	264	GLnGlnhIsglnAlaIglNhlhIsglnGlnSerhIshIsglnSerGlnhAsnIshAsn---	282

```

QY 520 AAGATTCTTCATCCCGGATGAAAGGTGAGCTCCCTTGCCGCCCTCAGCGAGCCAG 579
Db 283 -----Gln 283
QY 580 CGTGGGAGCACTCTCCCGGAGCAAGGC-----CAGGCCCTGAGGGGCACTTCTCAG 633
Db 284 HisArgAsnHisAsnGlnSerHisSerGlyProHisHisLeuProGlnAsnLeuMet 303
QY 634 GCCAGACAGATT-----GATTCCCGCTGGGAGATCTGTGTCCACCCAGCTTGTGTGT 687
Db 304 ProArgCysMetLeuValAspTyrProIleArgCysValValGlnGlyLysTyrHisAla 323
QY 688 GGCATCATCGGAAAGAGGCGCTTGACCATTAAGACATCATCAACAGACACCCAGCCCGG 747
Db 324 ValIleIleGlyProAsnGlySerThrIleLysAspIleAlaSerSerThrArgCysArg 343
QY 748 GTAGAT-----ATCCATAGAAAGAAAGAACTCT-----GAGCTGCAGAGAGACCT 792
Db 344 ValAspPheValAsnLeuSerLysGlnArgThrValLeuGlyAsnAsnAspArgIle 363
QY 793 GTCACCATTCATGCCACCCCGAGGGGACTTGTGAAGCATGCGCGCATGATTCTTGAATC 852
Db 364 LeuThrValHisGlyValAlaGlnGlnAlaThrLysAlaValAlaArgIleLeuAspVal 383
QY 853 ATGCAAGAAAGAGGCA-----GATGAGACCAAACTAGCCGAGAGATTCTCTGAATC 906
Db 384 IleGlnSerGlnAlaValLysAspAspValAsnValGlyAlaAspThrValLeuArgMet 403
QY 907 TTGGCACAACATGCTGTGGTGGAACTGATTGGAAGAAAGAGGCAAAATTTGAAGAA 966
Db 404 ArgAlaHisAsnGlnLeuGlySerGlyArgLeuIleGlyLysAlaGlySerSerLysGln 423
QY 967 ATTGAACATGAACAGGAGGCAAGATTAAGATCTGTCAGAGATTGAGCATATAC 1026
Db 424 IleMetGlnLysThrGlyThrAsnIleThrLysTyr-----IleGln 439
QY 1027 AACCCG-----GAAAGAAC 1041
Db 440 ProProGlyGlyIleSerGlyLeuThrAlaAsnGlnLeuGlyLeuMetGlnArgThr 459
QY 1042 ATCACTGTGAAGGC-----ACAGTTGAGCCTGTGCGACTGTGAGATAGATTAAG 1098
Db 460 IleMetValArgGlyProSerIleGlnAlaValAlaGlnAlaGlnAlaLeuIleSerAla 479
QY 1099 AAGCTCGCGAGGCGCTTGAATATGAT----- 1125
Db 480 LysLeuLysLysCysTyrGlnSerAspSerGlnLeuArgAlaGlnSerMetGlnCysPro 499
QY 1126 -----ATGCTGCTGTTTACACCCACTCCGATTAATCTTCCAGCCTGTATCCC 1173
Db 500 MetProProMetMetMetProProIleLeuProProGlyAlaSerSerSerAlaValSer 519
QY 1174 CATCACCAGTTTGCCCGTTCG-----CATATCACTTTTATCCAGAG 1218
Db 520 AlaProHisPheIleProThrProValGlyValMetGlnValGlnHisPheAlaSerSer 539
QY 1219 CAGGAGATTGTGAAT----- 1233
Db 540 GlnHisLeuValHisGlnAsnAlaAsnAsnSerPheLeuGlnProGlyValLeuGlnIle 559
QY 1234 -----CTCTCATCCCAACCCAGCGCTGTGGC 1260
Db 560 GlnProGlyThrThrAsnLeuArgGlnValArgMetTyrValProAspSerMetIleGly 579
QY 1261 GCCATCATCGGAAAGAGGCGGCAACATCAACAGTGCAGATTCCGCGAGCTCT 1320
Db 580 AlaLeuIleGlyLysGlyLysAsnIleLysMetIleLeuArgAspThrGlyAlaSer 599
QY 1321 ATCAAGATT-----GCCCTGCGAAGGCCAGAC----- 1350
Db 600 ValLysIleGlnAlaProGlnGlnLysThrGlnArgGlnAlaGlnGlnAlaGlnLysLys 619
QY 1350 ----- 1350

```

```

Db 620 ArgLysLeuAspGlnThrAspSerGlyCysGlnGlyValAlaSerGlyAspHisProGln 639
QY 1350 ----- 1350
Db 640 GlnPheLeuGlnAspAsnAlaThrIleAsnSerSerAspAlaIleGlnGlnLysProLys 659
QY 1351 ---GTCAAGCAAGAGATGATCATCATCAACCGGCGCCAGCGAAGCCAGTCAAGGCCAG 1407
Db 660 ProValSerGlnArgMetValThrIleAsnGlyAspAspLeuGlnLeuLysAlaGln 679
QY 1408 GAGCGGATCTTTGGGAAACTGAAGAGAGAAACTTTTAACCC----- 1452
Db 680 SerTyrValPheSerLysIleAlaGlnThrSerSerLeuProSerSerGlyMetAsp 699
QY 1453 -----AAGAAAGAGTGAAGCTGGAGAGCGCATATCAGATGCGCTCTTCCACAGCT 1503
Db 700 GlyAspArgSerHisMetLeuArgIleArgThrGlnValSerValProThrArgIleIle 719
QY 1504 GCGCGGATGATTTGGCAAGAGTGGCAAGACCGTGAACGAACTGCAGAACTTAACAGTGA 1563
Db 720 GlyArgIleIleGlyLysGlyGlyGlnAsnValArgGlnLeuGlnArgIleThrGlyAla 739
QY 1564 GAATCATCGTCGCTCGTGCACCAACGCCAGAT----- 1596
Db 740 ValValLysIleProGlnGlnGlnLysArgAsnGlyGlyGlnValTyrArgHisAspAspGly 759
QY 1597 ---GAAATGAGAGATGATCGTCAGAAATTAATCGGCACTTCTTGCAGCGAGCTGA 1653
Db 760 LeuGlnGlnAspMetThrMetIleArgThrIleGlyAsnMetLysSerThrHisAsnVal 779
QY 1654 CAGCGCAAGATCAAGGAAATTTTACACAGGTG-----AAGCAGCAG 1695
Db 780 GlnPheArgLeuAlaHisLeuValAsnGlnLysTyrTyrArgSerGlyAspHisArgAsnLys 799
QY 1696 GAGCAGAAATNCCCTCAGGAGTCCGCTGCAGCGGAGCAAGT 1738
Db 800 SerSerAspTyrLysGlyGly-ArgProHisSerAlaProSer 813

RESULT 2
T19216
Hypothetical protein C12D8.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T19216
R:McMurray, A.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19092
A:Accession: T19216
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-589 <MW>
A:Cross-references: EMBL:Z73969; PIDN:CAA98232.1; GSPDB:GN00023; CESP:C12D8.1a
A:Experimental source: clone C12D8
C:Genetics:
A:Gene: CESP:C12D8.1a
A:Map position: 5
A:Introns: 7/3; 25/1; 159/3; 318/3; 513/2

Alignment Scores:
Pred. No.: 8,49e-11 Length: 589
Score: 274.00 Matches: 118
Percent Similarity: 41.13% Conservative: 79
Best Local Similarity: 24.63% Mismatches: 183
Query Match: 4.66% Indels: 99
DB: 2 Gaps: 18

US-09-270-437D-8 (1-3283) x T19216 (1-589)
QY 667 GTCCCAACCAAGTTTGTGTGTCATCATCGAAGAGGCGCTTGACCATTAAGACAATC 726
Db 54 IleProGlnSerAlaValGlyIleValIleGlyArgGlyGlySerGlnIleGlnGlyIle 73

```



```

QY 997 ATCTCATCTTTCAGAT---TTGAGCATATACACCCGGAAGAACCATCATCTGTGAG 1053
Db 190 Leu-----ValGlnAspAsnGlnSerValSerAspGlnSerLysProLeuArgLleThr 207
QY 1054 GGCAAGATTGAGGCGCTGTGCCAGTGTGAGATTATGAAAGAAAGCTGTGAGGCGC 1113
Db 208 GlnAspProGln-----LysLleGlnLeuAlaLysGlnLeuValAlaGln 222
QY 1114 TTTGAAAATGATANGCTGCTGTTTACACCCATCCCGATACTTCTCCAGCTGTACCC 1173
Db 223 IleLeuAsnSerGlyGlyAspGlyAsnGlyLysGly-----LeuGlnMet 238
QY 1174 CATGACCAAGTTTGCGCGCTGCCGATCATCACTTTTCCAGACAGAGATGTGAT 1233
Db 239 HsHsAlaGlyGly-----GlyGlyAlaSerAlaArgGlyGlyValVal--- 254
QY 1234 CTCTTCATCCCAACCCAGCTGTGGCGCCCATCATCGGAGAGAGAGGAGCACATCAA 1293
Db 255 -----ValProArgSerSerValGlyLleLleLleGlyLysGlnGlyAspThrLleLys 272
QY 1294 CAGTGGCGAGATTGCGCGAGCTCTTATCAAGATTGCCCTTGGAGAGGCCAGACGTC 1353
Db 273 ArgLeuAlaMetGlnThrGlyThryLysLleGlnPheLysProAspAspProSerThr 292
QY 1354 AGCGAAAGATGTCATCATCAACCGGCGCACCGGAGGCCAGTTCAAGGCCAGGAGGAGCG 1413
Db 293 ProGlnArgCysAlaValLleMetGlyThrArgAspGlnLleTyrArgAlaThrGlnArg 312
QY 1414 ATCTTTGGAAAAGTGAAGAGGAAACTTCTTTAAC-----CCCAAA 1455
Db 313 IleThrGlnLeuValLysLysSerThrMetGlnGlnGlyGlyGlyAsnValAlaGly 332
QY 1456 GAAGAAGTGAAGCTGGAACG-----CATATCAGAGTGCCTCTTCCACAGCTGCG 1506
Db 333 AlaMetValSerAsnGlnLysLeuSerThrPheTyrMetSerValProAlaLysCysGly 352
QY 1507 CGGATGATTGGGAAGAGTGGCAAGCCGTAAGAGAACTGACAGAACTTAACAGTGCAGAA 1566
Db 353 LeuValLleGlyLysGlyLysGlnThrLleLysGlnLysSerGlnSerGlyAlaHis 372
QY 1567 GTCATGCTGCTCGTACCAACCGCCAGATGAATAAGAAAGTATGATCTGTC----- 1617
Db 373 CysGlnLeuSerArgAspProThrGlyAsnAlaAspGlnLysValPheValLleLysGly 392
QY 1618 -----AGAATTATCGGCACTTCTTGTAGCCACATGTCACAGCGCAAGATCAGGAA 1671
Db 393 GlyLysArgAlaLleLleLys-----AlaLysHisLeuLleArgLleLysValGlyAsp 410
QY 1672 ATTGTACACAGGTGAAGCAGCAGCAGCAAAATACCTCAGGAGTGCCTCAGCGCC 1731
Db 411 IleAlaProAsnThr----- 415
QY 1732 AGCAAGTGAAGCTCCACAGGCACACAGCAAAACAGATGAATGATAGCCCTTCCACAC 1791
Db 416 ----- 419
QY 1792 -----CTGACAGATGAGCAACAAAGCAG-----CCAGCCAGATG----- 1827
Db 420 AspSerAlaMetThrMet-----GlnThrGlnPheSerAlaProAlaGlnAsnAsnPhe 437
QY 1828 GAGAGCAACCAACAGCATCTGAGATGAGATGAGTCTGCGAGCGCGCGCAGGACTCTGC 1887
Db 438 GlyGlyGlnGlnGlnThrPheAsnProValAlaGlnLleProAlaAlaAlaGln----- 454
QY 1888 CGAGGCTGTGAGAACCCCAAGGAGCGGAGGAGGAGGAGTCAAGCTTTTGCCAG 1947
Db 455 -----AsnProTyrGlnValGlyGlyTyrGlnGlnAsnSerValTyrAlaGln 470
QY 1948 AACCAACGAGGCGCGCTCCCGCCCGCCAGGAGGCTTCTGCGAGGCTTCAAGCATCA 2002
Db 471 GlnThrAlaAlaProAlaAlaAlaProTyr-AlaAlaAlaGlyLleValGlnPro 488

```

```

T49962
hypothetical protein F8M21.160 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T49962
C:By: R. Bevan, M. Murphy, G. J. Riley, P. J. Hudson, S. Bancroft, I. J. Mewes, H. W. Rudd, S. J. Le
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24493
A:Accession: T49962
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-568 <BEV>
A:Cross-references: EMBL:AL353993; GSPDB:GN00063; ATSP:F8M21.160
A:Experimental source: cultivar Columbia; BAC clone F8M21
C:Genetics:
A:Gene: ATSP:F8M21.160
A:Map position: 5
A:Introns: 200/3; 337/3; 544/3

Alignment Scores:
Pred. No.: 1,15e-10 Length: 568
Score: 272.00 Matches: 129
Percent Similarity: 39.19% Conservative: 94
Best Local Similarity: 22.67% Mismatches: 235
Query Match: 4.62% Indels: 112
DB: Gaps: 20

US-09-270-437D-8 (1-3283) x T49962 (1-568)
QY 535 CCGATGTAAGAGTGAAGCTCCCTTCGCGCCCTCAGCGAGCCGAGCCAGTGGGAGCCACTCT 594
Db 3 ProAspHisArgMet-----SerProAspHisArgAspSerHisArgLysArgSer 19
QY 595 TCCCGGAGCAAGAGCGACGCCCTCGGAGCACTTCTCAGGCGAGA----- 639
Db 20 ArgProGlnSerAspTyrAspAspAsnGlyGlySerLysArgGlyTyrArgGlyAspAsp 39
QY 640 -----CAGATTGATTTCCGCTGCGGATCTGCTCCACCG 678
Db 40 ArgAspSerLeuValLleAspArgAspThrValPheArgTyrLeuLysProValLys 59
QY 679 TTTGTTGTCGATCATCTGCGAAGAGAGGCTTGACCATTAAGAACTCATAGCAGAGCC 738
Db 60 LysLleGlySerValLleGlyArgGlyGlyAspLleValLysGlnLeuArgAsnAspThr 79
QY 739 CAGTCCCGGTTGATATCATGAAAGAAAGAACTGTGAGTGCAGAGAAAGCTGTACCC 798
Db 80 ArgSerLysLleArgLle-----GlyGlnAlaLleProGlyCysAspGlnArgValLleThr 98
QY 799 ATCCATGCAACCCAGAGGAGGACT-----TTGAA 828
Db 99 IleTyrSerProSerAspGlnThrAsnAlaPheGlyAspGlyGlnLysValLeuSerPro 118
QY 829 GCATGCCGATGATTTTGAATCATGCAAGAAAGAGCAGATGAGCAACCAACTAGGCCAA 888
Db 119 AlaGlnAspAlaLeuPheArgLleHisAspArgValValAlaAspAspAlaArgSerGln 138
QY 889 GAGATTGCT-----CTGAATCTTGGCAACATGAGCTGTGTT 927
Db 139 AspSerProGlnGlyGlnLysGlnValThrAlaLysLeuLeuValProSerArgGlnLle 158
QY 928 GGAAGAATGATGGAAGAAAGAGCAGAAATTTGAAGAAATTTGAACATGAAACAGGGAGC 987
Db 159 GlyCysLleLeuGlyArgGlyGlyGlnLleValGlnAsnLleArgSerGlnThrGlyAla 178
QY 988 AAGATTAACAATCTCATCTTTCAGAGATTTGAGATA-----TACAACCGGAAAGAAC 1041
Db 179 GlnLleArgLleValLysAspArgAsnMetProLeuCysAlaLeuAsnSerAspGlnLeu 198
QY 1042 ATCATGTTGAAGGAGCAGTTGAGAGCTGTGCAAGTGTGAGATGATGATTAAGAGAG 1101
Db 199 IleGlnLleSerGlyGlnValLeuLleValLysAlaLeuLeuGlnLleAlaSerArg 218

```

```
QY 1102 CTGCGTGAAGGCTTTGAAAAATGATATGCTGCTGTTAAACCCACTCCGATATCTTCC 1161
    |||
    |||
Db 219 LeuHsgluAsnProSerArgSerGluAsnLeuSerSerSerGlyTyrProAla 238
QY 1162 AGCGCTGACCCCATCACCAGTTGGCCCG----- 1191
Db 239 GlySerLeuMetSerHisAlaGlyGlyProArgLeuValGlyLeuAlaProLeuMetGly 258
QY 1192 -----TTCGGCATCATCATCT 1209
Db 259 SerTyrGlyArgAspAlaGlyAspTyrSerArgProLeuTyrGlnProArgAsnAsp 278
QY 1210 TATCCAGACGACGAG--ATTGGAATCTCTTCAATCCCAACCCAGCTGTGGGCGCCATC 1266
    |||
    |||
Db 279 ProProAlaThrGluPhePheIleArgLeuValSerProValGluAsnIleAlaSerVal 298
QY 1267 ATCGGGAAGAAGGGGCGACATCAATCAACAGCTGGCGAGATTCCGCGAGCTTATCAAG 1326
    |||
    |||
Db 299 IleGlyTyrGlyGlyAlaLeuIleAsnGlnLeuArgGlnGlnIuThrArgAlaThrIleTys 318
QY 1327 ATT-----GCCCTCGGCGAAGGCCCGAC-----GTCAAGGAAGAAGATGTC 1368
    |||
    |||
Db 319 ValAspSerSerArgThrGluGlyAsnAspCysLeuIleThrIleSerAlaArgGluVal 338
QY 1369 ATC-----ATCACCGGCGCACCGGAACCCAGTTCAAGCCCGACGAGCGATCTTT 1419
    |||
    |||
Db 339 PheGluAspAlaTyrSerProThrIleGluAlaValMetArgLeuGlnProTyrCysSer 358
QY 1420 GGGAACTGGAAGAAGGAAACTTTTAAACCCCAAGAAGAAGTGAAGTGAAGCCCAT 1479
    |||
    |||
Db 359 AspTyrSerGluAlaGlyAspSerGlyLeu-----ValSerPheThrThrArg 373
QY 1480 ATCAGATGCGCCCTCTTCCACAGCTGCGCGGTGATTTGGCAAAAGTGGCAACCGTGAAC 1539
    |||
    |||
Db 374 LeuLeuValAlaProSerSerArgIleGlyCysIleLeuGlyTyrGlyAlaIleThr 393
QY 1540 GAACTGCAAGACTTAAACGAGTCAAGAGTC--ATCGTCTCTGTAACCAAGCCCA 1593
    |||
    |||
Db 394 GluMetArgArgMetThrIleAlaAsnIleArgIleLeuGlyTyrGluAsnLeuProTys 413
QY 1594 -----GATGAAATGAGAAAGTATGCTCAGATTTATCGGCACTTTT 1638
    |||
    |||
Db 414 ValAlaSerAspAspGluMetValGlnValAlaAsnPheMetValLeuLeuTyrPhe 433
QY 1639 GTAGCCAGACTGACACGCGCAAGATGAGGAATTTTACAACAGTGAAGACGACGAG 1698
    |||
    |||
Db 434 SerLeuGlnPheLeuLeuSerArgLeuArgTyrPheThrTyrSerValAsnSerAsnSer 453
QY 1699 CA-----GAAATACCTTCAGAGAGTCCCTCAACAGCGACGACGAGTCCCAACAGCA 1754
    |||
    |||
Db 454 AsnPheAspValGlnIleSerGlyGluLeuAspValAlaTyrGluAlaLeuIleGlnIle 473
QY 1755 CCGCAAAACACGATGATGTA----- 1778
    |||
    |||
Db 474 ThrSerArgLeuArgAlaAsnValPheAspArgGluGlyAlaValSerAlaLeuMetPro 493
QY 1779 GCCCTTCCA--ACACCTGACAGATGAGCAACCAAGCCACCGACCGATCGGAGCAAA 1835
    |||
    |||
Db 494 ValLeuProTyrValPro-----ValAlaProAspAlaGlyAsp-ArgPheAspTyrAs 511
QY 1836 CCAAAAGCATCTGAGGAATGAGAACTGCGAGGCGCGCCAGGACTGCGGAGCGCC 1895
    |||
    |||
Db 511 PserArgAspSerArgArgLeuGlu-----ArgGlyAs 522
QY 1896 TGGAAACCCCAAGGCGCGAGAGGGGCGGAAAGTTCAGCAGTTTGGCAAGAACCG 1955
    |||
    |||
Db 522 nProTyrProGlyGlyTyrGlySerSerGly-----ValSerAlaGluGlyTyr 538
QY 1956 AGCCCGGCTCCGCGCCCGCAGGCG 1980
    |||
    |||
Db 538 rSerProTyrGlyAlaProValGly 546
```

```
RESULT 5
D96554
Hypochemical protein F19C24.19 [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C/Accession: D96554
R/Theologis: A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso
Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.F.; Hughes, B.; Huizart, L.
Nature 408, 816-820, 2000
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A/Reference number: A86141; MUID:21016719; PMID:11130712
A/Accession: D96554
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-621 <STO>
A/Cross-references: GB:AE005173; NID:G11094762; PIDN:AA629695.1; GSPDB:GN00141
A/Genes: F19C24.19
A/Map position: 1
Alignment Scores:
Pred. No.: 4.56e-10 Length: 621
Score: 263.50 Matches: 129
Percent Similarity: 38.99% Conservative: 87
Best Local Similarity: 23.29% Mismatches: 209
Query Match: 4.48% Indels: 129
DB: Gaps: 25
US-09-270-437D-8 (1-3283) x D96554 (1-621)
QY 616 CCGGCGGCGCACTTCTCAAGCGACAGATGATTCCCGCTGGGATCTTGTCCTACC 675
    |||
    |||
Db 9 ProAlaThrThrAlaThrAlaIleGlu--SerValHisPheArgLeuLeuCysProAla 27
QY 676 CAGTTTGTGTCATCATCATCGAAGAAGAGGCGCTTGACCAATAAGACATCATCAACGAC 735
    |||
    |||
Db 28 ThrArgThrGlyAlaIleIleGlyGlyGlyGlySerAlaIleArgHisLeuGlnSerVal 47
QY 736 ACCGATCCCGGTA----- 750
    |||
    |||
Db 48 ThrGlySerTyrIleArgValIleAspAspIleProValProSerGluGluArgValVal 67
QY 751 -----GATATCCATTAAGAAAAGAAC 771
    |||
    |||
Db 68 LeuIleIleAlaProSerGlyTyrSylsLysAspGluSerAsnValCysAspSerGluAsn 87
QY 772 TCTGAGCTGACAGAAAGCTGTCAACCATCAAGCAACCCCAAGGGGACT----- 822
    |||
    |||
Db 88 ProGlySerGluGluProTyrGlnGlnIuTyrGlySerGlnCysAlaGlyThrSerGlyGly 107
QY 823 -----TCTGAAGCATGCCGATATTTGTAATCATGACGAAA-- 861
    |||
    |||
Db 108 AspAspGluGluAlaProSerSerAlaGlnMetAlaLeuLeuArgValPheGluArgIle 127
QY 862 -----GAGGCAGATGAGAACCAACTAGCCGAAGAGATT 894
    |||
    |||
Db 128 ValPheGlyAspAspAlaAlaThrValAspGlyAspGluLeuAspLysGlyGlnSerGlu 147
QY 895 CCTCTG--AAATCTTGACACAAATGCTTGGAAGAGTGAAGTGAAGAAAGAGC 951
    |||
    |||
Db 148 GlyLeuCysArgMetIleValArgGlyAsnGlnValAspTyrLeuMetSerTyrGlyGly 167
QY 952 AGAAATTTGAAGAAATTTGAATCACTGAACAGGACCAAGATTAACAATTCATCTTTCAG 1011
    |||
    |||
Db 168 LysMetIleGlnIleArgGluSerSerGlyAlaIleValArgIleSerSerThrAsp 187
QY 1012 GATTTC-----AGCATATCAACCCGGAAGAACATCATCATGTGAAGGCAAGTTGAG 1065
    |||
    |||
```

```

Db      188 GlnIleProProCysAlaPheProGlyAspValValIleGlnMetAsnGlyLysPheSer 207
QY      1066 GCCTGTGCGCACTGCTGAGATGAGATTATGAAAGAGCTGCGTGAAGCC----- 1113
Db      208 SerValIysAlaLeuLeuValThrAsnCysLeuGlnGlnSerGlyAlaProPro 227
QY      1114 -----TTGAAATAATGATATGCTGGCTGTTAAACACCCGACCGGATAC 1155
Db      228 ThrTrpAspGlyCysProPheProGlnProGlyTyProProGlnTyHisSerMetLeu 247
QY      1156 TTCTCCAGCTGTATACCCCATCACCAGTTTGCCCGTTCGCGCAT----- 1200
Db      248 TyrHisProGlnTrpAspHisProProProAsnProMetProGluAspValGlyProPhe 267
QY      1201 CATCACTCTTATCCAGACAGAGAGATTGTG--AATCTCTTATCCCAACCGAGCTGTG 1257
Db      268 AsnArgProValValGluGluGluValAlaPheArgLeuLeuCysProAlaAspIysVal 287
QY      1258 GCGCGCATCATCGGGAGAGAGGCGGCACATCAACAGCTGGCGAGATTGCGCGGAGCC 1317
Db      288 GlySerLeuIleGlyLysGlyGlyAlaValAlaValAlaLeuGlnAsnGlnSerGlyAla 307
QY      1318 TCTATCAAGATTGCG--CCTGCGGAGGCGCCAGCGTACGGAAGAGATGCTCATATC 1374
Db      308 SerIleLysValSerAspProThrHis-----AspSerGluGluArgIleIleValIle 325
QY      1375 ACCGGGCGCACCGGAAGCCAGTTCCAG-----GCCGAGGGA-----CGGATC 1416
Db      326 SerAlaArgGluAsnLeuGluArgThrHisSerIleAsnValGlnAspGlyValMetArgVal 345
QY      1417 TTTGGGAAATGAAAGAGAAACTTTTAAACCCCAAGAAAGAGATGAGCTGGAAGCG 1476
Db      346 HisAsnArgIleValGluIleGlyPhe-----GluProSerAlaAlaValAla 362
QY      1477 CATATCAAGTGCCTCTTCCACAGCTGCGCGGTGATTTGGAAGGTGGCAAGCCGTG 1536
Db      363 ArgLeuLeuValHisSerProTyrlIleGlyArgLeuLeuGlyLysGlyGlyHisSerLeuIle 382
QY      1533 AAGCAAGTGCAGAACTTAAACAGTGCAGAGATC--ATCGTGCTCGTGAACCAAGCCCA 1593
Db      383 SerGlnMetArgArgAlaThrIleLysSerIleAsnValPheAlaLysAspGlnAlaThr 402
QY      1594 -----GATGAAATGAGAGATGATCTGCAGAAATATGAGGCACTTTTGTACCCAG 1647
Db      403 LysIlyrGlnSerGlnHisAspGluIleValGlnValIleGlyAsnLeu-----Lys 419
QY      1648 ACTGCACAGCGCAAGATCAGGAATTTGTAACAAGGTGAAGCAGACAGAGAAATAC 1707
Db      420 ThrValGlnAspAlaLeuPheGlnIleLeuCysArgLeuArg-----GlnAlaMetPhe 437
QY      1708 CCTCAGGAGTGGCTTCCACAGCGCACAGAGTGGCTCCACAGGACAGCAAGCAAAACAC 1767
Db      438 Pro-Gly--ArgLeuProPheGlnGlyMetGlyGlyProProProPheMetGly- 455
QY      1768 GGATGATGTAGCCCTTCCAAACAACCTGACAGAAATGAGACCAACGACGACGACAGATCG 1827
Db      456 -----ProTyProGluProProProProPheGlyPro----- 466
QY      1828 GAGCAAAACCAAGACCATCTGAGAAATGAAAGTCTGCGGAGCGGCGGACGAGACTCTGC 1887
Db      467 -----ArgGlnTyProAlaSerProA 474
QY      1888 CGAGGCGCTGAGAAACCCGAGGCGCG--AGAGGCGGCGGAGAAAGTTCAGCAGAGTT 1941
Db      474 spArgTyHisSerProValGlyProPheHisGluArgHisCysHisGly--ProGlyP 493
QY      1942 TGCCAGAACCAACGAGGAGCCCGCC--TCCCGCCCCCAGGAGGCTTTCAGAGCTTACGCCA 1998
Db      493 heAspArgProProGlyProGlyPheAspArgProPro----- 505
QY      1999 TCACATTCACCATTCAGTGGATCTCTCTGAATCTCC 2036
Db      506 -----SerProMetSerTrpThrPro 512

```

```

RESULT 6
A53184
myc far upstream element-binding protein - human
N:Alternate names: FUSE-binding protein
M:Species: Homo sapiens (man)
C:Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text_change 05-Nov-1995
C:Accession: A53184
R:Duncan, R.; Bazar, L.; Michelotti, G.; Tomonaga, T.; Krutzsch, H.; Avigan, M.; Levens,
  Genes Dev. 8, 465-480, 1994
A:Title: A sequence-specific, single-strand binding protein activates the far upstream e-
A:Reference number: A53184; MUID:94170991; PMID:8125259
A:Accession: A53184
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-644 <DUN>
A:Cross-references: GB:U05040; NID:g460151; PID:AAA17976.1; PFD:g460152
C:Keywords: DNA binding

```

```

Alignment Scores:
Pred. No.: 5.4e-10 Length: 644
Score: 262.50 Matches: 140
Percent Similarity: 37.73% Conservative: 86
Best Local Similarity: 23.37% Mismatches: 237
Query Match: 4.46% Indels: 136
DB: 2 Gaps: 25

```

```

US-09-270-437D-8 (1-3283) x A53184 (1-644)

```

```

QY      336 CCCTCTCACTGCGAGTGGAGGTGTTGATGAGACTTTTGCTCAATATGAGACAGTGA 395
Db      10 ProSerSerGlySerLeuGlyGlyGlyGly-----GlyGlyGly 22
QY      396 GAATGTGAACAAGTCAACACAGACAGAGAAACCCCGTTGTCAAGTCAATATGCAAC 455
Db      23 GlyGlyGly-GlyValAsnAspAlaPheLysAspAlaLeuGlnArg-----AlaAr 39
QY      456 AAGAGAAAGCAAAATATGCCATGAGAGACTTAACCGGGCATCACTTGAAGACATGCTC 515
Db      39 GlnIleAlaIleLysIleGlyLysAlaGlyThrSerLeuAsnSerAsnAspIysG 59
QY      516 CTTCAAGATTTCTTCATCCCGGATGAGAGGTTGAGCTCCCTGCGCCCTCAGCGAGC 575
Db      59 YTrIleGlyGlnLysArgProLeuGlnAspLysAspGlnProAspAlaLysValAl 79
QY      576 CAGAGGATGATTTCCCGGAGCGAGAGCGCCCTGGGGGCACTTTCAGGC 635
Db      79 AlProGlnAsnAspSerPheGlyThrGlnLeu-----ProProMetHisGlnGlnIle 97
QY      636 CAGACAGATGATTTCCCGGAGATCTGTGCTGCCACCCAGTTTGTGGCCATCAT 695
Db      97 rArgSerVal--MetThrGlnGluIleLysValProAspGlyMetValGlyPheIleI 116
QY      696 CGGAAAGAGGGCTTGACATAAGAAACATCAAGACAGAGCCAGCTCCGGGTAGATAT 755
Db      116 eGlyArgGlyGlyGluGlnIleSerArgIleGlnGlnGlnSerGlyCysLysIleGlnI 136
QY      756 CGATGAAAGAGAACTTGAAGCTGAGCTGACAGAAAGCTGTACCATTCATGCCAAG 815
Db      136 e--AlaProAspSerGlyGlyLeuProGluArgSerCysMetLeuThrGlyThrProG 155
QY      816 GGGGACTTGTGAAGCTGCGCGCATGATTTTGAATATCAGCAAA- 861
Db      155 uSerValGlnSerAlaLysArgLeuLeuAspGlnIleValGlnLysGlyArgProAlaPr 175
QY      862 -----GAGGCGATGAGACCAAACTAGCCGAGGAGATTCTCTGAATCTTGGC 911
Db      175 oGlyPheHisHisGlyAspGlyProGlyAsnAlaValGln-----GluIleMetI 192
QY      912 ACACAAATGCGCTGTGTGAAGACTGATTTGAAAAGAGGAGAAATTTGAAGAAATTTGA 971
Db      192 eProAlaSerLysAlaGlyLeuValIleGlyLysGlyGlyGluThrIleLysGlnLeuG 212

```

```

QY 972 ACATGAAACAGGAGCAAGATACATCTTTCAGAGATTGAGCATATACACC 1031
    |||
    |||
    |||
Db 212 ngluarglaaglyVallysMetValMet-----llelnasp-----GlyPr 226
    |||
    |||
    |||
QY 1032 GGAAGAAGACC-----ATCACTGTGAAGGCGACAGTTGAGCGCTGTGCCAG 1076
    |||
    |||
    |||
Db 226 oglnmsnThrlyalaasplysProleuArglleThrlyAspProtyrlyValGlnl 246
    |||
    |||
    |||
QY 1077 TGCTGAGATAGAGATTATGAAGAAGCTGTCGAGCGCTTTGAAATGATATGCTGCT 1136
    |||
    |||
    |||
Db 246 nalalyselMetValleuGluLeuIleArgasp----- 257
    |||
    |||
    |||
QY 1137 TAAACACCCATCCGAGTACTTCTCCAGCTGTACCCCATCAACGATTGGCCGTTCC 1196
    |||
    |||
    |||
Db 258 -----GlnlylyPheArgGluVal-----Argsnlylyrly----- 269
    |||
    |||
    |||
QY 1197 GCATCATCTCTTATCCAGAGCAGAGATGTGATCTTTCATCCCAACCGCTGT 1256
    |||
    |||
    |||
Db 270 -----SerArglleGlylyAsnGlnlylyleaspyValProileProArgPheAlaVa 287
    |||
    |||
    |||
QY 1257 GAGCGCATCATCGGGAAGAAGGGGGCACATCAACAGCTGGCGAGATTGCGCGAGC 1316
    |||
    |||
    |||
Db 287 lGlyllevalllelyArgasnlylyGluMetllelysllyllelnasnapAlaGlyVa 307
    |||
    |||
    |||
QY 1317 CTCTATCAAGATTGCCCCCTGGGAAGGCCAGAGCTCAGCGAAGATGTGTCATCAC 1376
    |||
    |||
    |||
Db 307 lArglleGlnPheylsProAspArgly---ThrThrProIuArglleAlaGlnlleTh 326
    |||
    |||
    |||
QY 1377 CGGGCCACCGGAA-----GCCAGTTCAAGGCCAG----- 1407
    |||
    |||
    |||
Db 326 rGlyProProAspArgCysGlnHlsAlaAlaGlnlleleThrAspLeuArgSerVa 346
    |||
    |||
    |||
QY 1408 -----GAGCGATCTTGGGAA----- 1425
    |||
    |||
    |||
Db 346 lGlnAlaGlyAsnProGlylylyProGlylylyArglylyArglylyGlnl 366
    |||
    |||
    |||
QY 1426 -----CTGAAGAAGAAACTCTTTAAACCCCAAGA 1457
    |||
    |||
    |||
Db 366 yAsnTrpAsnMetGlyProProGlylylyleuGlnlylyPheAsnPhelle----- 382
    |||
    |||
    |||
QY 1458 AGAAGTGAAGCTGGAAGCGCATATCAGATGCGCTCTTCAACAGCTGGCGGTATGG 1517
    |||
    |||
    |||
Db 383 -----ValProthrglylyserThrlyleuIlellel 393
    |||
    |||
    |||
QY 1518 CAAGGTGCAAGACCGTGAAGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1577
    |||
    |||
    |||
Db 393 ylyrsglylylyGlnThrlyleuysertlleSerGlnInserGlyAlaArglleGlnleu 413
    |||
    |||
    |||
QY 1578 TCGTGAACAAACCGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1632
    |||
    |||
    |||
Db 413 nArgAsnProProProAsnAlaAspProAsnMetlyslleuPheThrlyleArglylyThr 433
    |||
    |||
    |||
QY 1633 -----TTCTTGTCTAGCCAGACTGCACAGCCGACAGATCAGGAAATTTGACACA 1682
    |||
    |||
    |||
Db 433 oglnGlnlleaspylyAlaArglleuIleGlnlylysllyleGlylyProValAsnPr 453
    |||
    |||
    |||
QY 1683 GGTGAAGCAGCAGAGAGAAATACCTCAGGAGTGCCTCAGAGCGCAGAGAGAG 1742
    |||
    |||
    |||
Db 453 oldeuGlyProProValProHlslylyProHlslylyAl-ProGlyProHlslylyProG 473
    |||
    |||
    |||
QY 1743 CTCGCAAGCAGCAGCAAAACCAAGATGATGATGATGATGATGATGATGATGATG 1802
    |||
    |||
    |||
Db 473 lyProProGly-ProGly-----ThrPro-----Met 481
    |||
    |||
    |||
QY 1803 AGACCAAGCAGCAGCAGCAGATCGGAGCAAAACCAAGACCATCTGAGGATGAGAAG 1862
    |||
    |||
    |||
Db 482 GlyProTyAsnProAlaProTyAsnProGlyPro----- 493
    |||
    |||
    |||
QY 1863 CTGCGAGGCGGCGAGGACTGCGGAGG-----CCCTGAGAACCCGAGGGCGGAGAGG 1919
    |||
    |||
    |||
Db 494 -----ProGlyProAlaProHlslylyProProAlaProTyAlaProGlnly 509
    |||
    |||
    |||
QY 1920 GGGGGGAAAGTCAAGCCAGGTTTGCAGAA-----CAACGAGCCCCCGC 1964
    |||
    |||
    |||

```

```

Db 510 TrpGlyAsnAlaTyrrProHlsTrpGlnGlnAlaProProAspProAla 526
    |||
    |||
    |||
RESULT 7
127855
hypothetical protein ZK418.9 - Caenorhabditis elegans
C|Species: Caenorhabditis elegans
C|Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
R|Fulton, U.
submitted to the EMBL Data Library, April 1994
A|Description: The sequence of C. elegans cosmid ZK418.
A|Reference number: Z20430
A|Accession: T27855
A|Status: preliminary; translated from GB/EMBL/DBJ
A|Molecule type: DNA
A|Residues: 1-768 <FNU>
A|Cross-references: EMBL:U00047; PIDN:AA50693.1; CESP:ZK418.9
A|Experimental source: strain Bristol N2
C|Genetics:
A|Gene: CESP:ZK418.9
A|Introns: 166/1; 221/3; 243/1; 279/1; 387/3; 490/1; 567/3; 727/2

Alignment Scores:
Pred. No.: 6,67e-10 Length: 768
Score: 261.50 Matches: 129
Percent Similarity: 39.92% Conservative: 83
Best Local Similarity: 24.29% Mismatches: 226
Query Match: 4.44% Indels: 93
DB: 2 Gaps: 20

US-09-270-437d-8 (1-3283) x T27855 (1-768)
QY 560 CGCCCCCTC--AGCGAGCCAGCGTGGGACCATCTTCCGGGAGCAAGCCAGCCC 616
    |||
    |||
    |||
Db 234 ArgProleuAspserGlnlleleuAspGlyAspLeuIleProthrylyslsSerSerGlu 253
    |||
    |||
    |||
QY 617 CTGGGGGCACTT--CTCAGCGCAACAGATTATTTCCGTCGGATCCTGTCCCA 673
    |||
    |||
    |||
Db 254 ValGlyAspLeuAsnMet-GlyAspSerAspLylyleThrAspIleTyrrPro-ValProG 273
    |||
    |||
    |||
QY 674 CCAAGTTGTGTGGTCATATCGCAAGAGAGAGGCTTGACCATTAAGAAATCACTAAG 733
    |||
    |||
    |||
Db 273 luyrvalValGlyleuVallleGlylyseGlyserGlnIleArgllelleGlnlnt 293
    |||
    |||
    |||
QY 734 AGACCAAGTCCCGGTAGATATCATAGAAAGAAAGAACTGTGAGCTGACAGAGAGCTG 793
    |||
    |||
    |||
Db 293 hrserGlyCysArgValGlnMetAspProAspHlsGlnserValAsnGlyPheArgAsnc 313
    |||
    |||
    |||
QY 794 TCACCATTCATGCCACCCCAAGAGGAGCTTTCAGAGCATGCCGATGATTTTGAATCA 853
    |||
    |||
    |||
Db 313 ysrThrIleGlnlylyProProAspGlnValAlaValAlaArgGlnMetlleThrGlnVal 333
    |||
    |||
    |||
QY 854 TG-----CAGAAAGAGCAGATGAGCAACCAATCAGCCGAAGAGATTCCTGAAA 904
    |||
    |||
    |||
Db 333 leAsnArgasnGlnThrlyAlaGlnProGlyAlaAlaProGlyGlnValThrGlnlun 353
    |||
    |||
    |||
QY 905 TCTTGCACACATGCTGTTGGAAGACTGATTGGAAGAAAGCAAGCAAAATTTGAGA 964
    |||
    |||
    |||
Db 353 etleuIleProAlaAspLylylleGlyleuVallleGlylysglylyGlnThrIleArg 373
    |||
    |||
    |||
QY 965 AAATTGAACATGAACAGGAGCCAGATTAACAATTCATCTTTCAGAGATTGAGCATAT 1024
    |||
    |||
    |||
Db 373 leValGlnGlnInserGly-----leuArgAsnCyAsnValVal 386
    |||
    |||
    |||
QY 1025 ACAACCCGGAAGAACATCATCTGTGAAGGCAACATTTGAGGCTGT-----G 1072
    |||
    |||
    |||
Db 386 al-----GlnGlnThrThrThrAlaThrGlyGlnProlyAspLeuArgMetlleGlyS 404
    |||
    |||
    |||
QY 1073 CCAAGTGTGAGATAGATTAAGAAAGAGCTGCGTGAAGGCTTTGGAATGATATCTGG 1132
    |||
    |||
    |||
Db 404 expProAlaAlaIleGlnThrAlaLyAlaIleuValHlsAsnIleMetAsn----- 420
    |||
    |||
    |||

```

QY	1133	CTGTTAACACCCAC--TCGGACTCTTCCAGCCCTGTAACCCCAATCACCAAGTTGGCC	118
		-----ASntrGinglYasnaLaProleuLeuglAargalAProHISglInProSerIleY	439
QY	421	-----	
		CGTTCCCGCAATCACTCTTTATCCAGAG--CAGAGATTGGAAATCTCTTATCCCA	1246
Db	439	lnpHeGlyglYglYrclYAlaaglInlunlaaglInlaYlsglygluValIleValProA	459
QY	1247	CCGAGCGCTGGGGCGCCATCATCGGGAAAGAGGGGGCACATTAACACAGCTGGCCGAGAT	1306
		rglueSerAlaGlyMetIleIleGlysglyglYgluMetIleYlsArgIuendLaLaG	479
Db	459		
QY	1307	TCGCGCGAGCCCTCATCAAGATTGCCCTCGGAGAGGCCCAAGCGTCAAGCAAGATGG	1366
Db	479	luthrGlyThrLysIleIleInPheYsProAspThrIraenProAsnSerGluAspArgIleA	499
QY	1367	TCATCATCACCCGGGCGACCCGGAAGCCAGTTCAAGGCCCAAGGACGGATCTTTGGGAAC	1426
Db	499	laValIleMetGlyThrArgAspGlnIleYrArgAlaThrGluArgIleThrGluIleV	519
QY	1427	TGAAGAAG-----GAAAATTCTTTTAACCCCAAGAAGAAGTGAACCTGGAACG----	1476
Db	519	alaenArgAlaIleYlsAsnAsnGlyAlaProGlnAspArgIlySerAlaGlyThrValL	539
QY	1477	-----CATATCAGAGCCCTCTTCCACAGCTGCCCGGGTGAATTG	1516
Db	539	eupProGlyInSerIlePheThyMetHISValProAlaGlyLysCysIlyeuValIleG	559
QY	1517	GCMAAGGTGGCAAGACCGGTGAACGAACTGCAGAACTTAACCAAGTGCAGAGATC--ATCG	1573
Db	559	lylYsglyglYgluAsnIleYsglnIleGluArgGluThrGlyAlaThrCysGlyIleUA	579
QY	1574	TGCCTCGTGCACCAAAAGCCCAATGAATGAAGAAATGATGTCAAGATTATCGGGCAC-	1632
Db	579	laProAlaAlaGluInlunYlsAsnGluAspGluYlsValPhe---GluIleYsglySerG	598
QY	1633	-----TTCTTCTAGCCACAGATGCACAGCGCGAATGCAGGAAATTGTACAAC	1681
Db	598	lnleuGlnIleHISHisAlaSerHISleuValArgIleYleValIleYlnIleSerProA	618
QY	1682	AGGTG-----AAGCAGCAGAGACGAAT	1705
Db	618	snThrProValProProleuGlnGlyAlaGlyIleYrIleGlnGlnGlnGlnAlaMetP	638
QY	1706	ACCTCAGGAGATC-----GGCTCAGAGCGGAGCAAGTGAAGGCTCCACAGAGCACAGC	1753
Db	638	heSerAlaGlyThrGlnAsnGlyglYrGlnSerThrGlyGlyPheIleGlnGlnGlnG	658
QY	1760	AAAAACAAGCATGAATGTAGCCCTTCCCAACACTGACAGATGAGACCAAAAGCACCGAG	1813
Db	658	lnGlnPro-----GlnPheGlnHis-----GlnGlnGlnGlnTrpA	670
QY	1820	CCAGATCGGAGCAAAACCAAGACCATCTGAGAGATGAGAACTGCGGAGCGGCAGAG	1875
Db	670	1a-----	670
QY	1880	GACTCTGCCGAGGCGCTTGAGAACCCCAAGGGCGCAGAGGGCGGGGAGATCAAGCAGG	1933
Db	682	roThrGluMetYrGlnAsnProMetGlnGlnProGlnAlaSerValIleProGlnAlaAs	702
QY	2000	CCACTTCACCATTCACACTGGATCTCCT	2028
Db	702	eralAspProAlaProAlaValAsnPro	711

RESULT 8
T25832
hypothetical protein M01A0.1 - *Caenorhabditis elegans*
C1:Species: *Caenorhabditis elegans*
C1Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T25832
R:Sheet: P.
Submitted to the EMBL Data Library, February 1997
A:Description: The sequence of C. elegans cosmid M01A10.
A:Reference number: Z20094
A:Accession: T25832
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-680 <CCH>
A:Cross-references: EMBL:U08174; PIDN:AA042272.1; GSPDB:GNO00019; CESP:M01A10.1
A:Experimental source: strain Bristol N2; clone M01A10
C:Genetics:
A:Gene: CESP:M01A10.1
A:Map position: 1
A:Introns: 63/2, 97/3, 121/1, 160/1, 269/3, 411/3, 512/1, 649/2

Alignment Scores:
Pred. NO.: 2.11e-09 Length: 680
Score: 254.00 Matches: 115
Percent Similarity: 40.26% Conservative: 100
Best Local Similarity: 21.54% Mismatches: 219
Query Match: 4.32% Indels: 100
DB: 2 Gaps: 17

US-09-270-437D-8 (1-3283) X T25832 (1-680)

[illegible]

QY 1045 ACTGTGAAGGCGACAGTGTGAGCGCTGTGCGACGTGCTGAGATGAGATTATGAAGAAGCTG 1104
 Db 292 GlnIleThnGlnProlys-----GlnValGlnHisAlaLys----- 304
 QY 1105 CGTGAAGCGCTTTGAAATGATATGCTGTGCTGTAAACCCACTCCGAGTACTCTCCACG 1164
 Db 305 -----AlaLeuValAlaAspIleLeu-----AspGlyPheAspGlu 316
 QY 1165 CTGTACCCCGCATCACACAGTTGGCCCGCTCCCGCATCATCACTCTTATCCAGAGAGAG 1224
 Db 317 CysProProAlaGlyMetAlaGlyAsnSerPro-----ValAlaAlaMet 331
 QY 1225 ATTGTGATCTCTTCATCCCAACCCAGAGCTGTGGCGGCATCATCGGGAAGAAGGGGCA 1284
 Db 332 SerLeuGlnValIysValProArgSerThrValGlyAlaIleMetGlyLeuGlnGlySer 351
 QY 1285 CACATCAAAACAGCTGGCGAGATTCGCCGAGACCTCTATCAAGATTGCGCCGAGAGGC 1344
 Db 352 AsnIleLysIleSerAsnGluThrGlnThrLysIleGlnPheMetProAspAsp 371
 QY 1345 CCAGACGTCAGCGAAGATGTGATCATCATCCGCGCCACCGGAAAGCCAGTTCAGAGCC 1404
 Db 372 ProLysLeuMetGluArgThrLeuValValIleGlyAsnLysAsnLysValIleValCys 391
 QY 1405 CAGGAGCGATCTTTGGGAACTGAAGAAGGAAACTTTTAAACCCCAAGAAGAGAGTG 1464
 Db 392 ---AlaArgLeuLeuGlnLysIleValGlnAlaAsnSerGluAsnAlaAsnThrProIle 410
 QY 1465 AAGCTGGAAGCGCATATCAAGATCCCTCTTCCACAGCTGGCCGGGATGTGCAAGGT 1524
 Db 411 SerLeu---PheTyrMetLeuIleProAlaSerLysCysGlyLeuValIleGlyArgGly 429
 QY 1525 GGCAAGACCGGGAAGACAGACAGACAGAACTTAAACAGTGCAGAGATCGTCCGTGAC 1584
 Db 430 GlyIleThrIleArgGlnIleAsnLysGlnSerGlyAlaTyrCysGlnMetSerArgAsp 449
 QY 1585 -----CAAAGCCGAGTGA 1599
 Db 450 ProSerIleSerAlaIleGlnLysGlnPheValIleArgGlySerGlnThrGlnValGln 469
 QY 1600 AATGAGGAAGTGTGTCAGAAATTATCGGCGACTTTCTTGCTAGCCAGACTGCACAGCGC 1659
 Db 470 HisAlaLysHisLeuIleArgVal----- 477
 QY 1660 AAGATCAGGGAAT-----GTAACAAGAGTGAAG 1689
 Db 478 LysValAlaLysPheProProAsnThrProTyrIleAsnThrArgAlaAlaGlnProLeu 497
 QY 1690 CAGCAGGAGCAGAAATACCTCCAGGAGTGGCGCTCAGAGCGAGCAGATGAGGTCCAC 1749
 Db 498 GlnPheSerHisGlnAsnProThrAlaIleAspSerTrpArgAlaGlnProPheThrThr 517
 QY 1750 AGGACACCGAAGAAACAACGATGATGATGACCTTCCACACCTGCAGAGATGAGACCA 1809
 Db 518 GlnHisGlnAsnSerLeuSerLeuProGlnProGlnAlaHisGlnPheProAsnLeuMet 537
 QY 1810 ACCGACGCCCGCATTCGGAGCAGAAACCAAGACATCTGAG 1851
 Db 538 AlaTyrSerAlaArgLeuGlyTyrGlnSerHisProProGln 551
 RESULT 9
 T41600
 probable pre-mRNA splicing factor - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C/Accession: T41600
 R/Lyne, M.; Rajadream, M.A.; Barrell, B.G.; Bothe, G.; Pohl, T.
 submitted to the EMBL Data Library, October 1998
 A:Reference number: Z22003
 A/Accession: T41600
 A/Status: Preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA

A:Residues: 1-398 <Lyn>
 A:Cross-reference: EMBL:AL031825; PIDN:CAA21234.1; GSPDB:GN00068; SPDB:SPCC757.09c
 A:Experimental source: strain 972h-; cosmid c757
 C:Genetics:
 A:Gene: SPDB:SPCC757.09c
 A:Map position: 3
 Alignment Scores:
 Pred. No.: 3.4e-09 Length: 398
 Score: 250.00 Matches: 97
 Percent Similarity: 39.74% Conservative: 56
 Best Local Similarity: 25.19% Mismatches: 150
 Query Match: 4.25% Indels: 82
 Db: 2 Gaps: 12
 US-09-270-437D-8 (1-3283) x T41600 (1-398)
 QY 505 GAGAACTACTCTTCATCAAGATTCTCTAC-----ATCCGAGTGAAGAGTGAAGTCC--- 555
 Db 16 GlnAsnSerPhePheAspValThrPheGlnAspGluProAspGlnThrThrSerThrAla 35
 QY 556 -----CCTTGCCGCCCT---CAGGAGCC 576
 Db 36 ThrGlyIleAlaLysValSerIleProThrProLysProSerThrProLeuSerThrLeu 55
 QY 577 CAGCGTGGGACCACTCTTCCCGGAGCAGACCGCCCTGGGAGCATTCCTCAGCC 636
 Db 56 ThrAsnLysSerThrIleGlnGlnSerMetThrAsnGlnProGluProThrSerGlnVal 75
 QY 637 AGACAGATT-----GATTTCCG 654
 Db 76 ProProIleSerAlaLysProProMetAspAlaThrTyrAlaThrGlnLeuThr 95
 QY 655 CTGGGATCCCTGGTCCCAACCAAGTTTGTGTGCATCATCGGAAAGAGGCTTGACC 714
 Db 96 LeuArgAlaLeuLeuSerThrArgGlnAlaGlyIleIleGlyLysAlaGlyLysAsn 115
 QY 715 ATAAAGACATCACTAAGACAGACCCAGTCCGGTGAATATCCATAGAAAAGAACTCT 774
 Db 116 ValAlaGlnLeuArgSerThrThrAsnValLysAlaGlyAla---ThrLysAlaValPro 134
 QY 775 GAGGTGCAGAGAAAGCTGTCCACCATCCATGCCACCCCAAGGGGACTTCTGAAGATGC 834
 Db 135 AsnValHisAspArgValLeuThrIleSerGlyProLeuGlnAsnValAlaArgAlaTyr 154
 QY 835 CGCATGATCTTGAATCATGCAGAAAGAGCA-----GATGAGACCAACTAGCC 885
 Db 155 ArgPheIleIleAspIlePheAlaLysAsnSerThrAsnProAspGlyThrProSerAsp 174
 QY 886 GAAAGATTCT-----CTGAAATCTTGGCAACAATGCTGTGTAAGACTGAT 939
 Db 175 AlaAsnThrProArgLysLeuArgLeuIleAlaHisSerLeuMetGlySerIleIle 194
 QY 940 GGAAGAAGAGCGGAAATTTGAAGAAATTTGAACATGAAACAGGACCAAGATACATC 999
 Db 195 GlyArgAsnLysLeuArgLysLeuIleGlnAspLysCysSerCysAspGlyIleAla 214
 QY 1000 TCATCTTGGAGATTGAGCATATACACCCGGAAGAAACCTCACTGTGAAGGGCACA 1059
 Db 215 Ser-----LysAspMetLeuProGlnSerThrGlnArgThrValGlnIleHisGlyThr 232
 QY 1060 GTTGAAGCTGTGCCAGTGTGATGATGATTAAGAAAGTGGCGTGAAGCTTTGAA 1119
 Db 233 ValAspAsnLeuHisAlaAlaIleTrpGlnIleGlyLysCysLeuIleAspAspTrpGln 252
 QY 1119 ----- 1119
 Db 253 ArgGlyAlaGlyThrValPheTyrAsnProValSerArgLeuThrGlnProLeuProSer 272
 QY 1120 -----AATGATATGCTGGCTGTAAACACCACTCC 1149
 Db 273 LeuAlaSerThrAlaSerProGlnGlnValSerProProAlaAlaProSerThrThrSer 292

Pred. No.: 4.19e-08 Length: 479
 Score: 234.50 Matches: 99
 Percent Similarity: 42.38% Conservative: 90
 Best Local Similarity: 22.20% Mismatches: 152
 Query Match: 3.98% Indels: 105
 DB: 2 Gaps: 19

US-09-270-437D-8 (1-3283) x C86275 (1-479)

QY 583 GGGGACCACTCTCCCGGAGAGAGCCAGCCCTCGGGGCGACTTTCAGCCAGAGC 642
 DB 10 GlyValArgSerAsnLeuGlnSerGlnPheThrGlyAsnGlyGlySerIysArgArgAsn 29
 QY 643 ATT-----GATTTCCCGCTGGGATCCTG 666
 DB 30 LeuHisAspGluThrAspGlnAsnValIleAlaSerGluAspThrValTyrArgTyrLeu 49
 QY 667 GTCCCGACCCAGTTTGTGGTCCATCATCGGAAAGAGGGCTTGACCATTAAGAACATC 726
 DB 50 CysProValLysLysThrGlySerIleIleGlyLysGlyGlyGlnIleAlaLysGlnIle 69
 QY 727 ACTAAGCAGACCCAGTCCCGGAGTATTCATAGAAAAAGAACTTGAGCTGCAGAG 786
 DB 70 ArgSerGluThrLysSerAsnMetArgIleAsn--GlnAlaLeuProGlyCysGluGln 88
 QY 787 AAGCTGTCCATCATCCAGCCAGAGGAGACTTCT 825
 DB 89 ArgValValThrMetTyrSerThrAsnGlnGluLeuAsnHisSpeGlyAspAspGlyGln 108
 QY 826 -----GAAGCATGCCGATGATTCCTTGAATTCAGAAAGAGCA 867
 DB 109 LeuValCysProAlaLeuAspAlaLeuPheLysValHisAspMetValValAlaAspAla 128
 QY 868 -----GATGACACCAACTAGCCGAAAG--ATTCTCTG 900
 DB 129 AspGlnAspAspGlyThrAspAspAspAsnAspLeuGlyGlyLysGlnThrValThrVal 148
 QY 901 AAATCTTGGCACAACATGCTGTGTAAGAGCATGTGAAAGAAAGAGAGAAATTG 960
 DB 149 ArgMetLeuValProSerAspGlnIleGlyCysValIleGlyLysGlyGlnValIle 168
 QY 961 AAGAAATTGAAACATGAACAGGAGCAAGATTAACAATCTCATCTTTGACAGATTGAGC 1020
 DB 169 GlnAsnLeuArgAsnAspThrAsnAlaGlnIleArgVal-----181
 QY 1021 ATATACACCCGGAAAGAACCATCATCTGTGAAGGGCACAGTTGAGCCTGTGCCAGTGT 1080
 DB 182 -----IleLysAspHisLeuProAlaCysAlaLeuThr 192
 QY 1081 -----GAGATGAGATTATG-----AAGAAGCTGCCGTGAGGCTTT 1116
 DB 193 LeuSerHisAspGlnLeuLeuIleIleGlyGlnProLeuValValArgGlnAlaLeu 212
 QY 1117 GAAATGATATGTGCTGTAAACCCACTCCGATACTTC-----1158
 DB 213 TyrGlnValAlaSerLeuLeuHisAspAsnProSerArgPheGlnHisLeuLeuLeuSer 232
 QY 1159 TCCAGCTGTACCCCATCACCAGTTGGCCCGCTTC-----CCG 1197
 DB 233 SerSerSerSerSerMetHisGlnProGlyAlaMetLeuMetSerAlaIleLeuThrSer 252
 QY 1198 CATCATCATCTTAT-----CCAGAGCAGGAGATTGTGAATCTC 1236
 DB 253 SerHisArgAsnTyrAlaValArgArgAspIleAlaAspAlaArgGluPheCysValCys 272
 QY 1237 TTCATC---CCAAACCAAGCTGTGGGCGCATCATCGGAAAGAGGGAGCAGCATCAAA 1293
 DB 273 PheIleCysProAlaGlnAsnValGlyGlyValIleIleLysGlyGlyGlyPheIleAsn 292
 QY 1294 CACCTGGCGAGATTGGCGGAGCTCTTATCAAGATTGCCCTGGGAAAGGCCAGACGTC 1353
 DB 293 GlnIleArgGlnGlnThrGlyAlaThrIleArgValAsnThrSerGluThrAspAsp---311

QY 1354 AGCGAAAGATGTCATATCATCAGCGGCGCACCGGAA-----1389
 DB 312 ---AspAspCysIleIlePheIleSerSerIysGlnPheTyrGluAspGlnSerProAla 330
 QY 1390 -----GCCCATGTAAGCCAGGAGCGAGATCTTTGGAAACAGT---AAGAGGAAAC 1440
 DB 331 ValAsnAlaAlaIleArgLeuGlnIleArgCysSerGluLysValGlyLysAspAlaAsn 350
 QY 1441 TTCCTTAAACCCCAAGAAAGTGAAGCTGGAAGGCGATATGAGTGCCTTCCTCCACA 1500
 DB 351 -----AspLeuAlaIleSerThrArgLeuLeuValSerSerSerGln 364
 QY 1501 GCTGGCCGGTGTGATTGGGAAAGGTGGCAAGACCGTGAACGATCGACAGACTTAACCACT 1560
 DB 365 IleGlyCysLeuIleGlyLysGlyGlyValAlaValIleSerGluMetArgSerValThrArg 384
 QY 1561 GCAGAAATC---ATCGTGCCTCGTGAACCAAGCCAGAT-----GAAATGAGGAAGTG 1611
 DB 385 AlaAsnIleArgIleLeuGlnIleGlnLysPylProLysIleAlaArgGluAspGlnGlu 404
 QY 1612 ATCGTCAGAAATTATCGGGCACTTCTTGTCTAGCAGACTGCAAGCAGCGAATCAGGGA 1671
 DB 405 MetValGlnIleThrGly-----SerProAspAlaAlaMetLysAlaLeuThrGln 421
 QY 1672 ATGTGACACAGGTGAAG 1689
 DB 422 ValIleLeuArgLeuArg 427
 RESULT 12
 S46109
 hNRNP complex protein homolog YBR233w - yeast (Saccharomyces cerevisiae)
 N/Alternate names: hypothetical protein YBR1531
 C/Species: Saccharomyces cerevisiae
 C/Date: 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 19-Apr-2002
 R/Dubois, B.; El Bakoury, M.; Glandsdorff, N.; Mesenguy, F.; Pierard, A.; Scherens, B.
 submitted to the Protein Sequence Database, August 1994
 A/Reference number: S45782
 A/Accession: S46109
 A/Molecule type: DNA
 A/Residues: 1-413 <DDB>
 A/Cross-references: EMBL:Z6101; NID:9536631; PIDN:CA85196.1; PID:9536633; MIPS:YBR233
 A/Experimental source: strain S286C
 C/Genetics:
 A/Gene: SGD:PPB2
 A/Cross-references: SGD:S0000437
 A/Map position: 2R
 C/keywords: nucleus
 Alignment Scores:
 Pred. No.: 8.87e-08 Length: 413
 Score: 229.50 Matches: 105
 Percent Similarity: 42.20% Conservative: 79
 Best Local Similarity: 24.08% Mismatches: 179
 Query Match: 3.90% Indels: 73
 DB: 2 Gaps: 19
 US-09-270-437D-8 (1-3283) x S46109 (1-413)
 QY 523 ATTTCCTCAACATCCCGGATGAAGAGTGAAGTCCCTTCGCCCTCAGGAGCCAGCCT 582
 DB 10 IleThrThrThrProThrThrValIleValIleValSerProAsnThrLeuLysArg--LysLys 28
 QY 583 GGGGACCACTCTCCCGGAGCA-----606
 DB 29 GlyGlnAspThrSerGlnGlnIleGlnAlaGlnIleLysArgValAlaLeuLysAsp 48
 QY 607 -----GGCCAGCCCTGGGGACACTTCTCAGGCCAGACAGATT-----GATTTCCG 654
 DB 49 AlaAspSerHisSerSerAspAsnAspHisAspSerProAspAsnValProSerAspValHis 68
 QY 655 CTGCGATCTCTGTCCTCCCAACCCAGTTTGTGTGTCATATCGGAAAGAGGCTTGACC 714

```

Db      69  LeuArgMetLeuCysLeuValIleuHisAlaSerLeuIleValIleGlnHisLeuGlyAlaIleThr 88
QY      715  ATAAAGAAATCATAGTAAAGCAAGCCAGTCCGGGAGTATGATCCATAGAAAAGAAACCT 774
Db      89  IleSerArgIleLeuSerGlnThrSerAlaArgIleAsnIleSerAsnAsnIleArg--- 107
QY      775  GAGGTGCAAGAAAGCTGTCACCATCCATGCGCAAGCCCAAGGGGAGCTTGTGAAGCATGCG 834
Db      108  GllValAlProGlnArgIleValIleValArgGlnThrCysAspArgValAlaIleAlaIleArg 127
QY      835  CGCATGATTTCTTGAATATCATGCAAGAAAGGCAAGATGAGACCAAACTAGCCGAA----- 888
Db      128  GlyMetIleValAlaArgAlaIleuGlnIleHisGlnIleAsnGlnIleAsnGlnIleGlnIle 147
QY      889  GAGATTTCTTGAATAATCTTGGCAACATGCGTGGTGTGAAGCACTGATTTGGAAGAAAGA 948
Db      148  GlnIleSerIleAsnLeuIleProHisIleMetGlyCysIleIleIleIleIleIleIleIle 167
QY      949  GGCAGAAATTTGCAAGAAATTTGAACATGAAACAGGAGCAACAGATTAACAATCTCATTTTG 1008
Db      168  GlySerArgLeuArgGlnIleGlnAspLeuSerAlaIleAlaIleuLeuPheAlaSerProAsn 187
QY      1009  CAGGATTTGACATATACACCCGAAAGAACATCATCTGTGAAGGCGACAGCTTGAAGCGC 1068
Db      188  Gln-----LeuLeuIleSerAsnAspArgIleLeuThrIleAsnGlyValProAspAla 205
QY      1069  TGTGCGCATGCTGAGATGAGATTTATGAAGAGCTGCGTGAAGCCTTGAATAATATATG 1128
Db      206  IleHisIleAlaThrPheThrIleSerGlnThrLeuAsn---PheGlnMetGlnSer 224
QY      1129  CTGGCTGTAAACACC---CACTCCGAGTACTTC-----TCCAGCCTG 1167
Db      225  ProGlnIleAsnValIleArgSerIleThrGlnProThrGlnPheAsnSerValIleu 244
QY      1168  TAACCCCATCACCAAGTTTGCCCGCTCCCGCAT-----CATCATCTTATTCAGAG 1218
Db      245  IleAspHisSerGlnProAsnThrIlePheHisGlnArgAsnHisGlnIleThrIleProSer 264
QY      1219  CAGGAGATTTGTC-----ATCTCTTCATCCCAACCAAGCTGCGGCGGCATC 1266
Db      265  AspIleLeuLeuSerIleThrIleProAsnIleAsnLeuProIleSerThr----- 281
QY      1267  ATCGGAGAAAGAGGGGCAACATCAAAACAGCTGGC-----AGATTCGCGGAGCCTCTT 1320
Db      282  -----LeuLeuSerMetAlaThrProGlnIleThrIleAlaSer 294
QY      1321  ATCAAGATTTCCCTCCGAAAGGCCCAAGC---GTACAGAAAGAGTGTCTATCATCAC 1377
Db      295  ValAlaAsnAlaThrAlaPheGlnProAsnPheValIleProAsnValIleThrValIleuAsp 314
QY      1378  GGGCCACCGGAAAGCCAGTTCAAGGCCCAAGGATCTTTGGAAACTGAAAGAGGAA 1437
Db      315  GlyProValIleIleSerProAlaProGlnIleAsnHisIleuLeuMetAsnIleValIleGlnGln 334
QY      1438  AACTTCTTTAAACCCCAAGAAAGAAAGTGAAGCTGAAAGCGCATTCAGAGTCCCTTCTCC 1497
Db      335  IlePheIleAspIleIle----- 340
QY      1498  ACAGCTGGCGGGGTGATTTGGCAAAAGGTGGCAAGACCGTGAACCAATTCGCAAGACTTAAC 1557
Db      341  PheValIleGlyAsnValIleGlyIleAspGlyIleHisIleAsnSerValIleGlnIleThr 360
QY      1558  AGTGCAGAAAGTATCGTGGCTCGTGAACAAAGCCAGATGAA---AATGAGAAAGTATC 1614
Db      361  GlyCysSerIleIleIle-----GlnAspProValGlnIleGlySerSerGlnIleArg 377
QY      1615  GTCAGAAATTTACGAGCACTTCTTGTAGCCAGACTGCA-----CAGCGCAAG 1662
Db      378  LeuThrIleArgGlyIleThrPheMetAlaSerGlnAlaAlaIleMetLeuIleSerAsnIle 397
QY      1663  ATCAGGAGAAATTTGTAACAACAGTGAAGCAGCAGAGCAAGAAATACCTT 1710
Db      398  Ile---GlnIleAspArgSerAsnAlaGlnIleArgIleArgSerPro 412

```

```

RESULT 13
S58529
Alpha-complex protein 1 - human
Alternate names: nucleic acid-binding protein; protein PCBP-1
C/Species: Homo sapiens (man)
C/Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 01-Dec-2000
C/Accession: S58529; S58523; S5678; S43489; S41378; S42472
R/Killedjan, M.; Wang, X.; Liebhafner, S.A.
EMBO J. 14, 4357-4364, 1995
A/Title: Identification of two KH domain proteins in the alpha-globin mRNA stability com
A/Reference number: S58523; MUID:96016208; PMID:7556077
A/Accession: S58529
A/Status: preliminary
A/Molecule type: protein
A/Residues: 1-356 <KID>
A/Accession: S58523
A/Status: preliminary
A/Molecule type: protein
A/Residues: 1-356 <KID>
A/Accession: S58523
A/Status: preliminary
A/Molecule type: protein
A/Residues: 125-139;251-265;315-323 <KID>
R/Lieffers, H.; Deigaard, K.; Celis, J.E.
Eur. J. Biochem. 230, 447-453, 1995
A/Title: Characterisation of two major cellular poly(UC)-binding human proteins, each co
A/Reference number: S5678; MUID:95331278; PMID:7607214
A/Accession: S5678
A/Status: nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1-204,'V',206-356 <LEF>
A/Cross-references: EMBL:X78137; NID:9460770; PIDN:CAA55016.1; PID:9460771
A/Experimental source: ANA cells (transformed human amnion cells)
A/Note: Submitted to the EMBL Data Library, March 1994
R/Asashkin, H.C.; Loukianova, T.; Degerdal, A.; Smeland, E.B.
Nucleic Acids Res. 22, 959-964, 1994
A/Title: Tissue specific expression and cDNA structure of a human transcript encoding a
A/Reference number: S43489; MUID:94203810; PMID:9152927
A/Accession: S43489
A/Status: translation not shown
A/Molecule type: mRNA
A/Residues: 1-298,'H' <AAS>
A/Cross-references: EMBL:Z29505; NID:9444020; PIDN:CAA82631.1; PID:9444021
C/Genetics:
A/Genes: GDB:HNRPX
A/Cross-references: GDB:344947
C/Keywords: RNA binding

Alignment Scores:
Pred. No.: 1.73e-07 Length: 356
Score: 225.00 Matches: 82
Percent Similarity: 40.40% Conservative: 61
Best Local Similarity: 23.16% Mismatches: 123
Query Match: 3.82% Indels: 88
DB: 2 Gaps: 11

US-09-270-437D-8 (1-3283) x S58529 (1-356)
QY      619  GGGGCACTTCTCAGCCAGACAGATTATTCGCGTGGCAATCTGTCGTCGACCCAG 678
Db      4  GlyValThrGlnSerGlyLeuSerValIleThrLeuThrIleArgLeuLeuMetHisGlyIle 23
QY      679  TTGTTGGTGCATCATCGGAAAGAGGGCTTGACCATTAAGAAATCATAGACAGACC 738
Db      24  GluValGlySerIleIleGlyIleGlyIleValIleArgIleValIleArgIleGlnIleThr 43
QY      739  CAGTCCGGGTATATTCATAGAAAAGAGACTGTGAGCTTCAGAGAAAGCTGTACCC 798
Db      44  GlyAlaArgIleAsnIle-----SerGlnGlyAsnIleProGlnIleIleThr 60
QY      799  ATCAATGCAACCCCAAGGGAGCTTCTGAAGCATGCGCATGATCTTGAATCATGAC 858
Db      61  LeuThrGlyProThrAsnAlaIlePheIleValAlaPheAlaMetIleIleAspIleGln 80
QY      859  AAGAGGCAATGAGACCAACTA-----GCCAAGAGATTCTCTG 900

```

```

Db      81 GluAspIleAsnSerSerMetThrAsnSerThrAlaIleAsnArgProProValThrLeu 100
QY      901 AAAATCTTGACACAAATGGCTTGGTAGAGACTGTGAAAGAGGCGAAATTTG 960
Db      101 ArgLeuValValProAlaThrGlnCysGlySerLeuIleGlyLysGlyCysLysIle 120
QY      961 AAGAAATTTGAACATGAAACAGGACCAAGATACAGATCTCATCTTTGACGATTGAGC 1020
Db      121 LysGluIleArgGluSerThrGlyAlaGlnValGlnValAlaGly-----AspMetLeu 138
QY      1021 ATATCAACCCGGAAAGAAACCATCTGCTGAAAGGCGACAGTTGAGCCCTTGCCAGTCT 1080
Db      139 ProAsnSerThrGluArgAlaIleThrIleAlaGlyValProGlnSerValThrGlnCys 158
QY      1081 GAGATGAGATT-----ATGAAGAAGCTGCGTGAAGGCGCTTTGAAATATATG 1128
Db      159 ValLysGlnIleCysLeuValMetLeuGluThrLeuSerGlnSerProGlnGlyArgVal 178
QY      1129 CTGGCTGTAAACACCCACTCCGATCTCTCCAGCTG----- 1167
Db      179 MetThrIleProTyrGlnProMetProAlaSerSerProValIleCysAlaGlyGln 198
QY      1168 -----TACCCCAT----- 1176
Db      199 AspArgCysSerAspAlaAlaGlyTyrProHisAlaThrHisAspLeuGluGlyProPro 218
QY      1177 -----CACCAAGTTGGCCCG----- 1191
Db      219 LeuAspAlaTyrSerIleGlnGlyGlnHisThrIleSerProLeuAspLeuAlaLysLeu 238
QY      1192 -----TTCCGCATCATCAC----- 1206
Db      239 AsnGlnValAlaArgGlnGlnSerHisSerHisMetMetHisGlyGlyThrGlyPheAla 258
QY      1207 -----TCTTATCCAGAG----- 1218
Db      259 GlyIleAspSerSerSerProGluValLysGlyTyrTrpAlaSerLeuAspAlaSerThr 278
QY      1219 CAGAGAGTTTGATCTCTTCATCCCAACCCAGGCTGTGGGCGCATTCATGCGGAAGAG 1278
Db      279 GlnThrThrHisGluLeuThrIleProAsnAsnLeuIleGlyCysIleIleGlyArgGln 298
QY      1279 GGGGACACATCAACACAGCTGGCGAGATTCCCGGAGCGCTCTATCAAGATTGCC---CCT 1335
Db      299 GlyAlaAsnIleAsnGlnIleArgGlnMetSerGlyAlaGlnIleLysIleAlaAsnPro 318
QY      1336 GCGGAAAGCCCAACGTCAGCGGAAAGATGCTCATCATCAGGCGCAACCGGAAGCCGAG 1395
Db      319 ValGlnGlySer-----SerGlyArgGlnValThrIleThrGlySerAlaIleSerIle 336
QY      1396 TTCAGAGCCCAAGGAGCGATCTTTGGGAAACTGAAAGAGAA 1437
Db      337 SerLeuAlaGlnTyrLeuIleAsnAlaArgLeuSerGln 350

```

RESULT 14

T04533

hypothetical protein F28J12.30 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 14-May-1999

C:Accession: T04533

R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; submitted to the Protein Sequence Database, February 1998

A:Reference number: Z15377

A:Accession: T04533

A:Molecule type: DNA

A:Residues: 1-846 <BEV>

A:Cross-references: EMBL:AL021710

A:Experimental source: cultivar Columbia; BAC clone F28J12

C:Genetics:

A:Map position: 4

A:Initons: 94/3; 161/3; 191/3; 232/3; 262/1; 287/1; 293/3; 300/3; 503/3; 675/3; 7

```

Alignment Scores:
Pred. No.: 2.62e-07      Length: 846
Score: 224.00           Matches: 117
Percent Similarity: 33.61%      Conservative: 87
Best Local Similarity: 19.28%    Mismatches: 198
Query Match: 3.81%      Indels: 205
DB: 2                      Gaps: 20

```

US-09-270-437D-8 (1-3283) x T04533 (1-846)

```

QY      370 CTTTGGCTCAATATGGGACAGTGGAGATGGAAACAGTCACAC----- 417
Db      268 LeuLeuAspSerTyrGly-----HisThrIleGlyValAsnThrAlaThrPheThr 284
QY      418 -----GACACAGAAACCGCGTTGTCAACGTC 444
Db      285 ArgLysGlySerGlyMetSerSerGlyLeuIleGluThrAspSerGluValLeuAsnSer 304
QY      445 ACATATGCAACAAGAGAAAGCAAAATAGCCATGAGAAAGCTAAGCGGCATCAGTTT 504
Db      305 ValTyrLeuThrMetValGluArgLys----- 313
QY      505 GAGAACTACTCTTCAGATTTCTTACATCCCGGATGAAGAGTGAGCTCCCTTGCGCC 564
Db      314 -----Lys 314
QY      565 CCTAGAGGAGCCACGCTGGGAGCACTTTCCGGGAGCAAGCCACGCCCTGGGGGC 624
Db      315 ArgLysGlnIleGlnArgAsnAsnSerGlnSerAsnArg----- 327
QY      625 ACTTCTGAGGCGACACAGATGATTCCCG-----CTG 657
Db      328 ---AsnGlnLysValArgGlyLeSerHisAspLysIleAsnArgAspGluLeuValValTyr 346
QY      658 CGGATCCCTGGTCCGCCACCCAGTTTGTGGTCCCATCGGAGAAAGAGGCTTGACATA 717
Db      347 ArgIleLeuCysProIleAspValValGlyValIleGlyLysSerGlyLysValIle 366
QY      718 AAGAACTACTTAAGACAGACCCAGTCCGGGTAGATATCCATAGAAAGAACTCGA 777
Db      367 AsnAlaIleArgHisAsnThrLysAlaLysIleLysValPhe---AspGlnLeuHisGly 385
QY      778 GCTGCAGAGAGGCTGTGACCATCATGCAACCCCAAGAGGAGCTTGAAGCATGCCGC 837
Db      386 CysSerGlnArgValIleThrIleTyrCysSerValLysGluLysGlnGlu----- 403
QY      838 ATGATTCTTGAATATCATGACAGAAAGAGGAGATGAGACCAACTAGCCGAAGATTCT 897
Db      404 -----IleGlyPheThrLysSerGluAsnGluProLeuCysCysAlaGlnAspAlaLeu 421
QY      898 CTGAAATCTTGGCACCAATGCGTTGTT----- 927
Db      422 LeuLysVal-----TyrAspAlaIleValAlaSerAspGluGluAsnThrLysThr 439
QY      928 -----GGA 930
Db      440 AsnValAspArgAspAsnLysGluCysArgLeuLeuValProPheSerGlnSerSer 459
QY      931 AGACTGATTGGAAGAGGACAGAAATTTTGAAGAAATTTGAACATGAACAGGAGCAAG 990
Db      460 SerLeuIleGlyLysAlaGlyGluAsnIleLysArgIleArgArgAlaGlnAlaSer 479
QY      991 ATTAACATCTCATCTTTGACAGATTGAGC----- 1020
Db      480 ValLysValValSer---LysAspValSerAspProSerHisValCysAlaMetGluTyr 498
QY      1020 ----- 1020
Db      499 AspAsnValValIleSerGlyLysProGlnSerValLysGlnAlaLeuPheAlaVal 518
QY      1021 -----ATATAC-----AACCGGAAGACATCATCTGTAAGGCGCAGTTGAG 1065
Db      519 SerAlaIleMetCysTyrLysIleAsnProArgGluAsnIleProLeuAspSerThrSerGln 538

```

```

QY 1066 GCCTGTGCCAGTGTGAGATGAGATTGAGAAAGCTGCGTGAAGCC----- 1113
Db 539 AspyAlProAlaAlaSerValIleValProSerAspLeuSerAsnSerValIleProGln 558
QY 1114 -----TTTGAAGAAAT----- 1122
Db 559 ThrGlyPheTyrSerAsnGlnAspHisIleLeuGlnGlnGlyAlaGlyValProSerTyr 578
QY 1123 ---GATATGCTGGCTGTTAAACCCACCTCCGATAC----- 1155
Db 579 PheAsnAlaLeuSerValSerAspPheGlnGlyTyrAlaGlnThrAlaAlaAsnProVal 598
QY 1156 -----TTTCCAGCTGTGACCCCATCATCAGCTTTGGCCCTCCCGATCATCATC 1206
Db 599 ProValPheAlaSerSerLeuProValThrHisGlyPheGly-----GlySer 614
QY 1207 TCTTATCCAGAGACAGAGATTGTGAATCTTCATCCCAACCCAGCTGTGGCCCATC 1266
Db 615 SerArgSerGlnGlnLeuValPheLeuValLeuCyProLeuCyAsnIleMetArgVal 634
QY 1267 ATCGGAGAGAGAGGGGACACATCAACAGCTGCGAGATTGCGCGAGCTCTTATCAAG 1326
Db 635 IleGlyLysGlyGlySerThrIleLysArgIleArgGlnAlaSerGlySerCysIleGln 654
QY 1327 ATTGCCCTCGGAGAGGCCACACGTCAGCGGAAAGATGTGATCATCATCAGCGGACACG 1386
Db 655 ValAsnAspSerArgThrLysCysGlyAspAspGlnCysValIleIleValThrAlaThr 674
QY 1387 GAAGCCCATGTTCAAGGCCAGGAGACGATCTT----- 1419
Db 675 GlnIleLeuPheCysCysLeuSerThrProPheValPheMetGlnSerProAspMet 694
QY 1420 -----GGAAACTGAAAGAGAAACTTCTTAACTCCCAAGAGAA 1461
Db 695 LysSerMetAlaValGlnAlaValIleLeuLeuGlnGlnGlyTyrIleAsnAspGlnAspAla 714
QY 1462 GTGAAGCTGGAAGCGCATATCAGAGTCCCTCTCCACAGCTGGCGGGTGAATGGCAGAA 1521
Db 715 GlnAsnValIleSerGlnLeuValSerLysValIleGlyCysValIleGlyLys 734
QY 1522 GGTGGCAGACCGGTGACAGAACTGAACTTAACGATGACAGAACTGATCGCTCGT 1581
Db 735 SerGlySerValIleLeuGlnIleArgLysArgThrAsnAlaAsnIleCysIleSerLys 754
QY 1582 GACCAAAACCCAGATGAAATGAGAAGTGATGTGCAATTAATCGGCGCATTTCTTGCT 1641
Db 755 GlyLysLysAspAsp-----LeuValGlnValSerGlyGlnVal----- 767
QY 1642 AGCCACATCTGCACAGCGCAGAAATGTCAGGAATGTACAAAGGTGAAAGCAGCAGGACAG 1701
Db 768 -----SerSerValArgAspAlaLeuIleGlnIle-ValLeuArgLeuHis 782
QY 1702 AAATACCTTCAGAGAGTCCCTCAAGCGCAGCAAGTGAAGGCTCCACAGGACCAACGAA 1761
Db 782 ggluaspvalleuglyAspLysAspSerValAlaIleThrArgLysProProla----- 799
QY 1762 AACACAGGATGAATGTAGC 1780
Db 800 -ArgThrAspAsnCysSer 805

RESULT 15
A44125
high density lipoprotein-binding protein, 110K - human
C:Species: Homo sapiens (man)
C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 24-Sep-1999
C:Accession: A44125
R:McKnight, G.L.; Reasoner, J.; Gilbert, T.; Sundquist, K.O.; Hokland, B.; McKernan, P.A.
J. Biol. Chem. 267, 12131-12141, 1992
A>Title: Cloning and expression of a cellular high density lipoprotein-binding protein c
A:Reference number: A44125; MIMD:92291094; EMD:1318310
A:Accession: A44125
A>Status: preliminary

```

```

A:Molecule type: mRNA
A:Residues: 1-1268 <MKG>
A:Cross-references: GB:M44098; GB:M83789; NID:G183891; PIDN:AAA35962.1; PID:G183892
A>Note: Sequence extracted from NCBI backbone (NCBIN:106862, NCBIPI:106863)
C:Superfamily: vlg11m

Alignment Scores:
Pred. No.: 3,46e-07 Length: 1268
Score: 223.00 Matches: 104
Percent Similarity: 38.79% Conservative: 69
Best Local Similarity: 23.32% Mismatches: 173
Query Match: 3.79% Indels: 100
DB: 2 Gaps: 18

US-09-270-437d-8 (1-3283) x A44125 (1-1268)

QY 460 GACAGAGCAAAATATGCCATGAGAGAGTGAAGCGGG----- 495
Db 445 GlnAspValAsnValAlaGlnGlnGlnIleGlnGlyMetValLysAspLeuIleAsnArg 434
QY 496 -----CATCAGTTTGAAGAACTAC----- 513
Db 435 MetAspTyrValGlnIleAsnIleAspHisLysPheHisArgHisLeuIleGlyLysSer 454
QY 514 TCTTCAAGATTCTCTAATCCCGGATGAA--GAGGTGAGCTCCCTTCGCCCTCCAG 570
Db 455 GlnAlaAsnIleAsnArgIleLysAspGlnTyrLysValSerValArgIleProAsp 474
QY 571 CGAGGCCACGCGGGGACACCTTCCCGGAGAGCAAGCGCCCTGGGGCACTTCT 630
Db 475 SerGlnLys-----SerAsnLeuIleArgIleGlnGlyAsp---ProGlnGlyValGln 491
QY 631 CAGGCCAGACAG-----ATTGATTTCCCGCTGCGG----- 660
Db 492 GlnAlaLysArgGlnLeuLeuGlnLeuAlaSerArgMetGlnAsnGlnArgThrLysAsp 511
QY 661 ATCTGTGTCACCAACCGAGTTGTGTGTCATCATTCGAAAGAGGCTTGACCAATAAAG 720
Db 512 LeuIleIleGlnGlnArgPheHisArgThrIleIleGlnGlnGlyGlnArgIleArg 531
QY 721 AACATCATTAACAGACACCGATCCCGGATGATGATCATAGAAAGAAAGAACTGTGAGCT 780
Db 532 GlnIle--ArgAspLysPheProGlnValIleIleAsnPheProAsp-----Prola 548
QY 781 GCAGAGAACCTGTGACCATCATGACCCAGAGGAGACTTCTGAAGATGCCGATG 840
Db 841 ATCTTGAATCATGACAGAAAGAGCAGATGAGACCAAACTAGCCGAAAGATTCCTCTG 900
QY 841 TyrMetGlnLysMetValAlaAspLeuValGlnAsnSerLysSerIleSerValProIle 588
Db 901 AAATCTTGCGACACAAATGGCTTGTGGAAGCTGATGGAAAGAGCAGCAAAATTG 960
QY 961 PheLysGlnPheHisLys-----AsnIleIleLysGlyGlyAlaAsnIle 604
Db 961 AACAAATTAACAATGAACAGGACCAAGATTAACATCTCATCTTTCAGATTGAGC 1020
QY 1021 AATTAACACCCG--GAAGAACCATCATCTGTAAGGC-----ACAGTTGAGGCC 1068
Db 619 AlaGlnAsnSerAsnSerGlnThrIleIleIleThrGlyLysArgAlaAsnCysGlnAla 638
QY 1069 TGTGCCAGTGCATGATTAAGATTAAAGAAAGCTGTGAGGCTTTGAAATGATATG 1128
Db 639 AlaArgSerArgIleLeuSerIleGlnLysAspLeuAlaAsnIleAlaGln----- 655
QY 1129 CTGGCTGTTAAACCAACCATCCCGATACTTCTCCAGCCTGTACCCCAACCAAGTTTGGC 1188
Db 655 ----- 655
QY 1189 CCGTTCGGCATCATCATCTTATTCAGACAGAGAGATTGGAATCTTTCATCCCAACC 1248

```

```

Db      656 -----ValGluValSerIleProAla 662
QY      1249 CAGGCTGTGGCGCCATCATCGGGAAGAGGGGCACACATCAACAGCTGGCGAGATTG 1308
Db      663 LysLeuHisAsnSerLeuIleGlyThrLysGlyArgLeuIleArgSerIleMetGluGlu 682
QY      1309 GCCGAGCGCTCTATCAAGATTGCCCCGTGGGAAGGCCAGACGTACAGCGAAMAGATGTG 1368
Db      683 CysGlyGlyValHisIleHisPheProValGluGlySer-----GlySerAspThrVal 700
QY      1369 ATCATCACCGGAGCCACCGAAGCCCAAGTTCAGGCCAGGACGATCTTTGGAAACTG 1428
Db      701 ValIleArgGlyProSerSerAspValGluLysAlaLysLysGlnLeu--HisLeu 719
QY      1429 AAAGAGAAACTTCTTTAACCCCAAGAGAAGTGAAGCTGGAAAGCCATATCAGAGTG 1488
Db      720 AlaGluGlu-----LysGlnThrLysSerPheThrValAspIleArgAla 734
QY      1489 CCTCTTTCACAGCTGCGCGGTGATGTGCAAGGTGGCAAGACCGTGAACGACTGCGAG 1548
Db      735 LysProGluTyrHisLysPheLeuIleGlyLysGlyGlyLysIleArgLysValArg 754
QY      1549 AACTTAACCAAGTGCAGAGTATCGTGCCTCGTGACCAACCGCAGATGAATAATGAGGAA 1608
Db      755 AspSerThrGlyAlaArgValIlePheProAlaAlaGluAspLysAspGlnAsp----- 772
QY      1609 GTGATCGTCAGAAATTATGGGCACCTTTTGTCTAGCCGACTGCACAGCGCAAGATCAGG 1668
Db      773 --LeuIleThrIleIleGlyLysGluAspAlaValArgIleAlaGlnLysGlnLeuGlu 791
QY      1669 GAATTGTACCAACAGGTG 1686
Db      792 AlaLeuIleGlnAsnLeu 797

```

Search completed: July 23, 2004, 11:18:18
 Job time : 80.5997 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 23, 2004, 10:51:55 ; Search time 28.8809 Seconds

(without alignments)
11838.037 Million cell updates/sec

Title: US-09-270-437D-8

Perfect score: 5886

Sequence: 1 ggcacgagagagagcagaga.....aaccttgtaaatgttatttc 3283

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 11681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p model -DEV=x1p
-O=/cm2_1/USPRO.spool.p/US9270437/runat.23072004.095739.27413/app.query.fasta_1.5582
-DB=SwissProt.42 -QFMT=fastan -SUFFIX=isp -MINMATCH=0.1 -IOOPL=0 -IOOEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USPR=US09270437 @CGN 1.1 57 @runat.23072004.095739.27413 -NCPU=6 -ICPU=3
-NO MMAP -LARGEIOBERRY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	279	4.7	572	1	FUB3_HUMAN
2	277.5	4.7	721	1	FUB2_RAT
3	261	4.4	643	1	FUB1_HUMAN
4	259.5	4.4	651	1	FUB1_MOUSE
5	259	4.4	707	1	FUB2_HUMAN
6	259	4.4	769	1	ZBP2_CHICK
7	241.5	4.1	606	1	Y475_ARATH
8	240	4.1	339	1	PCB3_MOUSE
9	234	4.0	339	1	PCB3_HUMAN
10	229.5	3.9	403	1	PCB4_HUMAN
11	229.5	3.9	413	1	PBP2_YEAST
12	225	3.8	356	1	PCB1_HUMAN
13	225	3.8	356	1	PCB1_MOUSE
14	223	3.8	1268	1	VGAI_HUMAN
15	219.5	3.7	474	1	NOA1_RAT
16	219	3.7	403	1	PCB4_MOUSE
17	213	3.6	492	1	NOA2_HUMAN
18	211.5	3.6	362	1	PCB2_MOUSE

19	210	3.6	365	1	PCB2_HUMAN	Q15366	homo sapien
20	207.5	3.5	493	1	NOA1_MOUSE	Q9jkn6	mus musculus
21	207	3.5	510	1	NOA1_HUMAN	P51513	homo sapien
22	207	3.5	1270	1	VELN_CHICK	P81021	gallus gall
23	187.5	3.2	606	1	TDRH_HUMAN	Q9y2w6	homo sapien
24	184	3.1	629	1	PAB2_ARATH	P42731	arabidopsis
25	180	3.1	1222	1	S160_YEAST	P06105	saccharomyc
26	173	2.9	463	1	ROK_HUMAN	Q07244	homo sapien
27	173	2.9	463	1	ROK_MOUSE	O19049	oryzotolagus
28	172	2.9	660	1	YH11_EBV	P03181	epstein-bar
29	170.5	2.9	381	1	YBD2_YEAST	P38199	saccharomyc
30	167	2.8	5147	1	PCLD_HUMAN	Q9y6v0	homo sapien
31	166	2.8	2842	1	APC_RAT	P70478	rattus norv
32	160.5	2.7	470	1	NR54_HUMAN	Q15233	homo sapien
33	160.5	2.7	1781	1	AKI2_HUMAN	Q02952	homo sapien
34	157	2.7	1850	1	VIT2_CHICK	P02845	gallus gall
35	156	2.6	1943	1	PC15_MOUSE	O099p1	mus musculus
36	153.5	2.6	5262	1	MLT2_HUMAN	O14686	homo sapien
37	152.5	2.6	653	1	PABP_SCHPO	P31209	schizosach
38	152	2.6	892	1	ATX7_HUMAN	O15265	homo sapien
39	150.5	2.6	572	1	Y062_DEIRA	Q9yrm6	deinococcus
40	149.5	2.5	867	1	ATX7_MOUSE	O87411	mus musculus
41	149.5	2.5	3664	1	ELV3_HUMAN	Q14576	homo sapien
42	148.5	2.5	367	1	ELV3_MOUSE	O14576	homo sapien
43	148.5	2.5	511	1	YB87_BACTN	Q88290	bacteroides
44	148	2.5	1080	1	MRB1_MOUSE	P59759	mus musculus
45	147	2.5	784	1	YAV2_XANCV	P14728	xanthomonas

ALIGNMENTS

RESULT 1
FUB3_HUMAN
ID FUB3_HUMAN STANDARD; PRT; 572 AA.
AC Q96124; Q92346; Q95VB6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Far upstream element-binding protein 3 (FUSE binding protein 3).
GN FUBP3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=97094955; PubMed=8940189;
RA Davis-Smyth T., Duncan R.C., Zheng T., Michelotti G., Levens D.;
RT "the far upstream element-binding proteins comprise an ancient family
of single-strand DNA-binding transactivators.";
RL J. Biol. Chem. 271:31679-31687(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX TISSUE=Cervix, and Placenta;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marnusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uebis T.B., Toshlyuk S., Carninci P., Prange C.,
RA Kohn S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Pahay J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley A.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length

RT human and mouse cDNA sequences." U.S. Pat. Acad. Sci. U.S.A. 99:16699-16903(2002).
 CC Proc. Natl. Acad. Sci. U.S.A. 99:16699-16903(2002).
 CC -1- FUNCTION: May interact with single-stranded DNA from the far-
 CC upstream element (FUSE). May activate gene expression.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q96124-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q96124-2; Sequence=VSP_008323, VSP_008324;
 CC Note=No experimental confirmation available.
 CC -1- TISSUE SPECIFICITY: Detected in a number of cell lines.
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC
 CC EMBL: U69127; AAC50893.1; ALT FRAME.
 CC EMBL: BC001325; AAH01325.1; ALT_FRAME.
 CC EMBL: BC007874; AAH07874.1; --
 CC Genew, HGNC:4005; FDBP3.
 CC MIM: 603536; --
 CC InterPro: IPR004087; KH_dom.
 CC InterPro: IPR004088; KH_type_1.
 CC Pfam: PF00013; KH; 3.
 CC SMART: SM00322; KH; 3.
 CC PROSITE: PS50084; KH TYPE 1; 4.
 CC Transcription regulation; Trans-acting factor; Nuclear protein;
 CC DNA-binding; Repeat; Alternative splicing.
 CC KW DOMAIN 77 141 KH 1.
 CC KW DOMAIN 162 228 KH 2.
 CC FT DOMAIN 253 317 KH 3.
 CC FT DOMAIN 354 421 KH 4.
 CC FT VARSPLIC 1 64
 CC MAEIVGASAPVGMKAEEFVDAIHRVROLIAKIDSIPHLNN
 CC STPIYDPSYGVGVQKRPIDGV -> MPI (in
 CC isoform 2).
 CC /FTId=VSP_008323.
 CC Missing (in isoform 2).
 CC /FTId=VSP_008324.
 CC V -> D (in RBF 1).
 CC FT CONFLICT 72 72
 CC SEQUENCE 572 AA; 61640 MW; FIBB225542BC197D CRC64;
 SQ
 Alignment Scores:
 Pred. No.: 2,13e-10 Length: 572
 Score: 279.00 Matches: 117
 Percent Similarity: 39.96% Conservative: 82
 Best Local Similarity: 23.49% Mismatches: 202
 Query Match: 4.74% Indels: 98
 DB: 1 Gaps: 16
 US-09-270-437D-8 (1-3283) x PUB3_HUMAN (1-572)
 QY CATCACTTGAGAACTACTCCCTTCCATCTTCCTCATCCCGGATGAAGGTAGCTCC 555
 Db |||::: ||:: ||||| ||||| :::::
 Db HHSARGVALARGINILEALALALYSILESPSERILEPRO-----HLSLEAANAN 41
 QY CTTGACCCCTCCAGGACCCAGCGCTG-----GGACCACTCTTC 597
 Db ::|||::: ||:: ||||| ||||| |||||
 Db 42 SerThr-ProLeuValAspProSerValTYRGIYTYGIVAGLNLYSARGPROLEUAS 61
 QY CGGAGACAAAGCCAGCCGCTCCGGGGGACCATTCACAGCCAGACATGATGATCCGGCTG 657
 Db |||::: ||:: ||||| ||||| ||||| |||||
 Db 598 CGGAGACAAAGCCAGCCGCTCCGGGGGACCATTCACAGCCAGACATGATGATGATGATG 657
 Db |||::: ||:: ||||| ||||| ||||| |||||
 Db 61 pASpGIYALIGYASmGLINDELGYALAEUVALIHSGINARGTHRVALLIEHRLG-- 80

QY	658	CGGATCTGTCCTCCCAACCCAGTTGTTGTGGTCATCATCGGAAGAAGGGCGTTGACCAAT	717
Db	81	GluphelysValProaspysMetValGlyPheIleIleGlyArgGlyGlyGluGlnIle	100
QY	718	AAGAACCTCACTAAGGAGACCCAGTCCCGGATGATATCATATGAGAAAGAACTCTGGA	777
Db	101	SerArgIleGlnIleGluSerGlyCysIysIleGlnIle--AlaSerGluSerSerGly	119
QY	778	GCTGCAGAGAAGCCTGTGCACCATTCATGCCACCCGAGGGAGCATTTGAAACATGCCG	837
Db	120	IleProGluArgProCysValIleThrGlyThrProGluSerIleGluGlnIleAlaIysArg	139
QY	838	ATGATTCTTGAATTCATGACAGAA-----GAGCGAGAT	870
Db	140	LeuLeuGlyGlnIleValaIlePheArgCysArgAsnGlyProGlyPheHisAsnSpIleasp	159
QY	871	GAGACCAACTAGCCGAGAGATTCTCTGAAAATCTTGGCACACATGGCTTGTTGGA	930
Db	160	SerAsnSerThrIleGlnGlu-----IleLeuIleProAlaSerIysValGly	175
QY	931	AGACTGATTGTAAGAAAGAGACAGAAATTGAGAAATGATGAATGAATGAACAGGAGCCAG	990
Db	176	LeuValIleGlyArgGlyGlyGluThrIleLysGlnIleGlnGluArgThrGlyValIys	195
QY	991	ATAACAATCATCTATCTTTGGACAGATTGGACATA--TACACCCGGAGAAAGACCATCACT	1047
Db	196	MetValMet-----IleGlnAspGlyProLeuProThrGlyAlaAspIysProLeuArg	213
QY	1048	GTGAAGGGCACAGTTGAGGCGCTGTGCCAGTCGAGATGAGATTGTAAGAAAGCTCCGT	1107
Db	214	IleThrGlyAspAlaPheLysValGlnGlnAlaArgGluMetValIleGluIleIleArg	233
QY	1108	GAGGCTTTGAAATGATATGCTGGCTGT-----AACACCACTCCGGATAC	1155
Db	234	GluIysAspGlnAlaAspPheArgGlyValArgGlyAspPheAsnSerArgMetGlyGly	253
QY	1156	TTTCTCAGCCTGTGACCCCATCACCAAGTTGGCCCGTCCCGCATCACTCTTATCCA	1215
Db	254	GlySer-----	255
QY	1216	GAGCAGAGATTTGAAATCTTTCATCCCAACCCAGGCTGTGGCGCATCATCGGAG	1275
Db	256	-----IleGluValSerValProArgPheAlaValGlyIleValIleGlyArg	271
QY	1276	AAGGGGACACATCAAAACAGCTGGCGAGATTCCGCGAGCCTCTATCAAGATTGCCCT	1335
Db	272	AsnGlyGluMetIleLysIleIleGlnAsnAspAlaGlyValArgIleGlnPheLysPro	291
QY	1336	GCGGAAGGCCACAGCTCAGCGAAGAGTGTCATCAACCGGGCCACCGGAAGCCAG	1395
Db	292	AspAspGlyIleSer--ProGluArgAlaAlaGlnValMetGlyProProAsp--Arg	309
QY	1396	TTCAAGGCCACGAGCGAGCTTTGGAAACTG-----AAAGAGAAACTTC	1443
Db	310	CysGlnHisAlaAlaHisIleIleIleSerGluLeuIleLeuThrAlaGlnGluArgAspGly	329
QY	1444	TTTAACCCCAAGAAGAGTGAAG--	1467
Db	330	PheGlyGlyLeuAlaAlaIleArgGlyValArgGlyValArgGlyValAspIlePheValGly	349
QY	1468	-----CTGAAGCGCATATCAAGTCCCTCTTCCACAGCTGCGCGGGTG	1512
Db	350	AlaProGlyGlyValGlnGluIleThrIleThrValProAlaAspIysCysGlyLeuVal	369
QY	1513	ATTGCGAAAGGTGGCAACCGTCAGACGATCGCAAACTTAACACAGTCGAGAGTATC	1572
Db	370	IleGlyLysGlyGlyGluAsnIleLysSerIleAsnGlnIleSerGlyAlaHisValGlu	389
QY	1573	GTGCTCGGACCAAGCCAGATGAATGAATGAGAAAGATCGTCAGAAATTATCGGGAC	1632
Db	390	LeuGlnArgAsnProProProAsnSerAspProAsnIleu-----	402

```

QY 1633 TTCTTTGTAAGCCAGCTGCACAGCGCAAGATCAGGGAATTGTACAGAGGTGAACGAG 1692
DB 403 -----ArgArpPheThrIleArgGlyValProGlnGlnIleGluVal 416
QY 1693 CAGGAGCAGAAATATCCC-----TAGGAGTGCCTCTCAGACCGCAG----- 1733
DB 417 A1aArgGlnLeuIleAspGluValGlyGlyThrAsnLeuGlyValAspGluAlaPhe 436
QY 1734 ---CAAGTGAAGCTCCACAGCAGCAGCAAAACAGATGAATGTAGCCCTTCCACA 1790
DB 437 GlyGlnSerProPheSerGlnProProAlaProProHisGlnAsnThrPheProProAla 456
QY 1791 CCTGACAGAAATGACCAACAGCCAGCCAGCTGAGGAGCAACCA 1838
DB 457 SerSerGlyCysPheProAsnMetAlaAlaValAsnGlyAsnPro 472

RESULT 2
FUB2_RAT
ID_FUB2_RAT STANDARD; PRT; 721 AA.
AC Q99PF5;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Far upstream element binding protein 2 (FUSP binding protein 2) (KH
DE type splicing regulatory protein) (KSRP) (MAP2 RNA trans-acting
DE protein 1) (MARTAL).
GN FUBP2 OR KHSRP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A., SEQUENCE OF 73-87; 89-108 AND 475-486, SUBCELLULAR
RP LOCATION, AND FUNCTION.
RC TISSUE=Brain.
RX MEDLINE=22246918; PubMed=12358751;
RA Rehbein M., Wege K., Buck F., Schweizer M., Richter D., Kindler S.;
RT "Molecular characterization of MARTAL, a protein interacting with the
RT dendritic targeting element of MAP2 mRNAs.";
RL J. Neurochem. 82:1039-1046(2002).
CC -1- FUNCTION: Part of a ternary complex that binds to the downstream
CC control sequence (DCS) of the pre-mRNA. Mediates exon inclusion in
CC transcripts that are subject to tissue-specific alternative
CC splicing. May interact with single-stranded DNA from the far-
CC upstream element (FUSE). May activate gene expression (by
CC similarity). Binds to the dendritic targeting element and may play
CC a role in mRNA trafficking.
CC -1- SUBUNIT: Part of a ternary complex containing FUBP2, PTBP1, PTBP2
CC and HNRPH1 (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear. A small proportion is also found in
CC the cytoplasm of neuronal cell bodies and dendrites.
CC -1- SIMILARITY: Contains 4 KH domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AF308818; AAG59811.1; -.
CC InterPro: IPR004087; KH dom.
CC DR InterPro: IPR004088; KH_type_1.
CC DR Pfam; PF00013; KH; 4.
CC DR SMART; SM00322; KH; 4.
CC DR POSITIVE; PSS0084; KH TYPE 1; 4.
CC KM Transport; mRNA transport; mRNA processing; mRNA splicing;
CC Transcription regulation; Trans-acting factor; Nuclear protein;
CC DNA-binding; RNA-binding; Repeat.
FT DOMAIN 145 209 KH 1.
FT FT DOMAIN 234 300 KH 2.
FT FT DOMAIN 323 387 KH 3.

```

```

FT DOMAIN 425 492 KH 4.
FT REPEAT 572 685 4 X 12 AA IMPERFECT REPEATS.
FT REPEAT 572 583 1.
FT REPEAT 618 629 2.
FT REPEAT 644 655 3.
FT REPEAT 674 685 4.
FT DOMAIN 7 68 GLY/PRO-RICH.
FT FT DOMAIN 69 498 GLY-RICH.
FT FT DOMAIN 499 613 ALA/GLY/PRO-RICH.
SQ SEQUENCE 721 AA; 74226 MW; 482C7A765C60EE4A CRC64;

Alignment Scores:
Pred. No.: 2,84e-10 Length: 721
Score: 277.50 Matches: 161
Percent Similarity: 34.08% Conservative: 81
Best Local Similarity: 22.68% Mismatches: 283
Query Match: 4.71% Indels: 187
DB: 1 Gaps: 27

US-09-270-437d-8 (1-3283) x FUB2_RAT (1-721)
QY 104 CGCGCGTCACCGCGCAGCAGCTCCGCGCAGCTCTTGGGACAGAGAGCTGCCGCGG 163
DB 10 ProProGlyProProProProProAlaGlyGlyGly-GlyGlyAlaAlaGlyAlaGlyGly 29
QY 164 GACAGGTCTGCTGAAGTCCGCGCTACGCGCTTGTGAGTACCCGACGAGAACTGGGCA 223
DB 29 yProProGlyProProGlyProProGlyAlaGlyAspArgGlyGlyGlyProGlyGlyGly 49
QY 224 TCCGGGCGCATGAGACCCCTCTCGGGTAAAGTGAAATTCATGGGAAATCATGAAAGTTG 283
DB 49 yPro-----GlyGlyGlyGlyAla----- 55
QY 284 ATTACTAGTCTTAAAGGCTAAAGAGAGCAAGAAATTCAGATTGAAACATCCCTCTC 343
DB 56 -----SerGlyGlyProSerGly 61
QY 344 ACCTCAGTGGAGGTGTTGATGACTTTG-----GCTCAATATGAGCAGTGGAGA 397
DB 61 nProProGlyGlyGlyGly-ProGlyIleArgGlyAspAlaPheAlaAspAlaValGln 81
QY 398 ATGTGAACA-----GTCACACAGACACAGAAACCGCGTTGTCAAGTCA 445
DB 81 rgAlaArgGlnIleAlaAlaValIleGlyAlaPalaAlaThrValAsnAsn 101
QY 446 CATATGCAACAAGAGAAAGCAAAATATGACATGAGAGAGCTAAGCGGATGATTG 505
DB 101 hr-----P 102
QY 506 AGAATCACTCTTCAAGATTCTTCAATCCCGATGAGAGGTAGCTCCCTTGCCCC 565
DB 102 roAspPheGlyPheGlyGlyGlnIlyAspGlnIleGlnIleAspGlyAspGlnIleProAspSer 122
QY 566 CTGACGAGCGCGCGGCGGACCATCTTCCGCGGAGCAAGC-----CAGCGCCCTG 619
DB 122 ysisIleuAlaSerGlnGlyAspSerIleGlySerGlnIleGlyProIleHisProPro 142
QY 620 GGGGCACTTCTCAGGCGCAGACAGATTGATTCGCGTGGCGGATCTGTGCCACCCAGT 679
DB 142 rokrghrIserMetThrGlnGlu-----TyrArgValProAspGly 156
QY 680 TTGTTGGTCATCATCGAAGGCTTGACCATTAAGAAACATCATCACTAAGACAGACC 739
DB 156 etValGlyLeuIleIleGlyArgGlyGlyGlnIleAsnIleGlnIleAspSerG 176
QY 740 AGTCCGGGTGAGTATTCATAGAAAGAGAACTGTGAGCTGACAGAGAAGCTTCA 799
DB 176 lyCysIlyValGlnIle---SerProAspSerGlyGlyLeuProGlnIlySerValSer 195
QY 800 TCCATGCAACCCGACGAGGAGCTTGTGAACATGATCGCGATGATTTGAATATCATGAGA 859
DB 195 eurthrGlyAlaProGlnIlySerValGlnIlySerMetIleuAspAspIleValSer 215

```

```

QY 860 AA-----GAGCAGATGAGCAACAAAC 880
Db 215 rgglyarvglygylProProglylnPheHlsasPsnalAsnlglylnHsnlgly 235
QY 881 TAGCCGAGAGATTCTCTGAAATCTTGCGACACATGCTGTGGTGAAGACTGATTG 940
Db 235 hrvalglnhlllmetlleProAla-----GlylAslaGlyleuValllleg 251
QY 941 GAAAGAGAGCGAATTTGAAGAAATTTGAACATGAAACAGGAGCCAGATACATC- 999
Db 251 lylvglygylglnhrlllelysglnleuglnlgualarglaGlyalllyseuTlleuI 271
QY 1000 -----TCAATCTTGAGCATTTGACATATACACCCGAAAGAACATCACTGTGAAG 1054
Db 271 leglnasPglyserGlnsnhrsnValAsplysProleuArglllellleglyAspProt 291
QY 1055 GCACAGTTGAGGCTGTGCGACGTGCTGAGATAGAGATTATGAAAGACTGCTGAGGCTT 1114
Db 291 yrlvsValGlnlglnlaCysglumetValMetAspIleleuArg----- 305
QY 1115 TTGAAATGATATGCTGCTGTTAACACCCACTCCGATCTTCTCCAGCTGTACCCCT 1174
Db 306 --GluArgAspGln----- 309
QY 1175 ATCACCAGTTTGCCCGTCCCGCATCATCACTTTATCCAGAGCAG-----GAGATTG 1228
Db 310 -----GlyllyPhegllyAspArgasnlglylryGlyserArgValGlylglyl 326
QY 1229 TGAATCTCTTATCCACCCAGGCTGTGGGCGCATCATCGGAGAAAGGGGCGACACA 1288
Db 326 leaspValProValProArgHlsSerValGlyValValllleGlyArgserGlyglumetI 346
QY 1289 TGAACAGCTGGCGAGATTGCGCGGAGCCTCATAGATTGCCCTGGGAGAGGCCGAC 1348
Db 346 lelylsylleGlnsnAspArglyValArglllelnhelysglnAspAspGlylthng 366
QY 1349 AGCTGAGCAAGAGATGTCATCATACCCGCGCACCGGAGCCCATCAAGGCCCGACG 1408
Db 366 ly---ProgllylrystleAlahlsilemeGlyProProAsp---ArgCysgluHlsAlaA 384
QY 1409 GACGATCTTTGGGAACTGAAAGAGAA----- 1437
Db 384 lArArglleleAsnAspIleuGlnSerleuArgserGlyProProglyProProglyA 404
QY 1438 -----AACTCTTTAAC 1450
Db 404 lAProglymetProProglylyArglylArglylArglylnlglyAsnTTPglyProP 424
QY 1451 CCAGAGAGAGAGAGAGTGAAGCGCATATCAGATGCCCTTCCAGAGCTGCGCGG 1510
Db 424 roglygllylmetcthrPheSer-----lleProthrHlslyscysGlyleuV 440
QY 1511 TGAATGCAAGGTGCGACAGCCGTGAACGACTGACGAACTTAAACCAAGTGCAGAACTCA 1570
Db 440 allleGlylrygllygllyuAsnVallylAlaalleAsnlglnhtrGlylAlaPheValG 460
QY 1571 TCGTGCCTGTGACCAACCCCA-----GATGAATATGAGAGAGATGCTG----- 1617
Db 460 lnlleSerArglnleuProProAsnGlyAspProsnPheIysleuPheValllleArg 480
QY 1618 -----AGATTTATTCGGGCACTTCTTGCTAGCCAGACTGACAGCGCAGATCAGAG 1669
Db 480 lyserProglnglnllleAspHls-----AlalysGlnleuIleGlnlglyllleGlu 498
QY 1670 AAAT---TGACACACAGTGAAGCAGAGCAGAGCAGAAATACCTTCAGGGAG----- 1717
Db 498 lyProleuCysProValGlyProglyProglyProglyProAlaGlyPrometGlyP 518
QY 1718 -----TGCGCTCACAGCGCAGCA- 1735
Db 518 robHlsProglyProPheAsnlglnlyProProglylAlaProProHlsAlaGlylryP 538
QY 1736 -----AGTGAAGGCTCCACAGGACACAGCAAAACAGGATGATGATGCCCTT 1784

```

```

Db 538 roProProHlsGlnlyrProProglnglylryPgllyAsnThrGlyrProgluProPro- 557
QY 1785 CCAGACCTGACAGAAATGAGCAACAGCAGCCAGATCGGAGCAACCAAGAC 1844
Db 558 ProAlaProHls-----AspProsnlyslAlaAlaAlaAlaThrAspProHlsAla 575
QY 1845 ATCTGAGA-----ATGAGAAgTCTGCGAGCGCGCCAGGAGCTCTGCGAGGCCC 1895
Db 576 AlArtpAlaAlaTyryTyrySerHlsTyryGlnlglnProProglyProValPro----- 593
QY 1896 TGAGAACCCCGAGGCGCGAGAGGCGGAGAGGTCAAGCAGTTTGCAGAACACCG 1955
Db 594 -----GlyProAlaProAlaProAlaAlaProProAlaGlnlglylryProPro 609
QY 1956 AGCCCGGCTCCGCCCCCAG 1977
Db 610 Gln-ProProProThrGlylryGln 616

RESULT 3
FUBI_HUMAN STANDARD; PRT; 643 AA.
ID FUBI_HUMAN Q96AE4; Q12828;
AC 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Far upstream element binding protein 1 (FUBP binding protein 1) (FBP)
DE (DNA helicase V) (HDH V).
GN FUBP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1); SEQUENCE OF 191-193; 203-206;
RP 272-279; 284-290; 300-314; 321-328; 394-397; 409-411; 430-438 AND
RP 440-443; AND FUNCTION.
RC TISSUE=Leukemia;
RX MEDLINE=94170991; PubMed=8125259;
RA Duncan R., Bazari L., Michelotti G., Tomonaga T., Krutzsch H.,
RA Avigan M., Levens D.;
RT "A sequence-specific, single-strand binding protein activates the far
RT upstream element of c-myc and defines a new DNA-binding motif.";
RL Genes Dev. 8:465-480(1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Cassavani T.L., Schreiber T.E.,
RA Brownstein M.U., Ustin T.B., Toshlyuk S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulhaly S.J.,
RA Bosak S.A., McMan P.J., McKernan K.J., Malek U.A., Gunaratne P.R.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Faley J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 45-64; 133-145; 271-283; 308-321; 379-386; 414-424 AND
RP 430-439.
RX MEDLINE=21127960; PubMed=11222755;
RA Vindigni A., Ocham A., Triolo G., Falaechi A.;

```



```

CC      Name=2;
CC      Isoid=Q91WU8-2; Sequence=VSP_008322;
CC      Note=No experimental confirmation available;
CC      -1- PTM: Ubiquitinated. This targets the protein for proteasome-
CC      mediated degradation (by similarity).
CC      -1- SIMILARITY: Contains 4 KH domains.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; BC014763; AAH14763.1; -
DR      EMBL; AK029458; BAC26457.1; -
DR      MGD; MGI:1926294; D3E1D330E.
DR      MGD; MGI:1926294; 9530027K12R1K.
DR      InterPro; IPR008160; Collagen.
DR      InterPro; IPR004087; KH dom.
DR      Pfam; PF00013; KH; 4.
DR      SMART; SM00322; KH; 4.
DR      PROSITE; PS50084; KH_TYPE_1; 4.
KW      Transcription regulation; Trans-acting factor; Nuclear protein;
KW      DNA-binding; Repeat; Ubl conjugation; Alternative splicing.
FT      DOMAIN 96 160 KH 1.
FT      DOMAIN 181 247 KH 2.
FT      DOMAIN 271 335 KH 3.
FT      DOMAIN 372 439 KH 4.
FT      DOMAIN 16 21 POLY-GLY.
FT      DOMAIN 345 392 GLY-RICH.
FT      DOMAIN 446 556 PRO-RICH.
FT      VARSPLIC 67 67 D->DSGWTNPSSTTHWEGMSPSPFKD (in isoform
FT      2).
FT      CONFLICT 93 93 /FTID=VSP_008322.
FT      SQ SEQUENCE 651 AA; 68539 MW; F4EECA62FD9FA0D5 CRC64;
Alignment Scores:
Pred. No.: 4.01e-09
Score: 259.50
Percent Similarity: 36.50%
Beet Local Similarity: 23.24%
Query Match: 4.41%
DB: 1
Gaps: 27
US-09-270-437d-8 (1-3283) x FUBI_MOUSE (1-651)
QY      336 CCCTCTCACTCACTGAGGAGGTGTTGATGAGACTTTTGGCTCAATATGGACAGTGA 395
      |||||  ::::|  |||||  |||||  |||||
      10 ProseSerGlySerIaGlyGlyGly-Gly----- 20
QY      396 GAATGTGAACAAGTCAACACAGACAGAAACCGCGTTGTCAGCTCATAT----- 450
      |||||  |||||  |||||  |||||  |||||
      21 -----GlyValValAsnAspAlaPheIysAs 29
QY      451 -----GACACAGACAGACAGAAATATGCGATGAGAGAGTGAAGCGGACA 497
      |||||  |||||  |||||  |||||  |||||
      29 PalaLeuGlnArgAlaArgGlnIleAlaIaIaIysIleGlyGlyAspAlaGlyThrsIle 49
      ::::|  ::::|  ::::|  ::::|  ::::|
      49 uAsnSerAsnAspIlyrGlyrGlyrGlyrGlyrGlyrGlyrGlyrGlyrGlyrGlyr 69
      ::::|  ::::|  ::::|  ::::|  ::::|
      558 TTCG-----CCCGCTCAGCGAGCCAGCGTGGGACCACTTCCCGGAGACA 605
      |||||  |||||  |||||  |||||  |||||
      69 cAspAlaIysIysValProProGlnAsnAspSerPheGlyAlaGlnLeu----- 85
QY      606 AAGGACAGCGCCCTGGGGGACATTTCTAGCGCAGACAGATGATTTCCCGCTCGGATCCT 665
      |||||  |||||  |||||  |||||  |||||
      86 -----ProPomethIsgInGlnIleSerArgSerVal---MetThrGlnGluIlyrI 102

```

```

QY      666 GGTCCCAACCAAGTTTGTGTCATCATGGAAGAGGGCTTGACCATTAAGAACAT 725
      |||||  |||||  |||||  |||||  |||||
      102 sValProAspGlyMetValGlyPheIleIleGlyArgGlyGlyGlnIleSerArgI 122
      ::::|  ::::|  ::::|  ::::|  ::::|
      726 CACTAAGAGACCCAGTCCCGGTAGATATCCATAGAAAAGAACTGTGAGCTGCACA 785
      ::::|  ::::|  ::::|  ::::|  ::::|
      122 eGlnGlnIleSerGlyCysIysIleGlnIle---AlaProAspSerGlyGlyLeuProG 141
      ::::|  ::::|  ::::|  ::::|  ::::|
      786 GAAGCTGTGACCATCCATGCGACCCAGAGGGAGCTTCTGAGCATGCCGATGATCT 845
      ::::|  ::::|  ::::|  ::::|  ::::|
      141 uArgSerCysMetLeuThrGlyThrProGlnSerValGlnSerAlaIysArgLeuAs 161
      ::::|  ::::|  ::::|  ::::|  ::::|
      846 TGAATCATGCAGAA-----GAGCAGATGAGACCAACT 881
      ::::|  ::::|  ::::|  ::::|  ::::|
      161 pGlnIleValGlnIlyrGlyrProAlaProGlyPheHisIleGlyAspGlyrProGlyAs 181
      ::::|  ::::|  ::::|  ::::|  ::::|
      882 AGCGAAGACAGATTCCTGAAAATCTTGACACACAAATGGCTTGGAGACTGATGG 941
      |||||  |||||  |||||  |||||  |||||
      181 nAlaValGln-----GlnIleMetIleProAlaSerIysAlaGlyLeuValIleG 198
      ::::|  ::::|  ::::|  ::::|  ::::|
      942 AAAAGAGGACAGAAATTTGAAGAAATGAACATGAACAGGACCAAGATTAACATCTC 1001
      |||||  |||||  |||||  |||||  |||||
      198 YLysGlyGlyGlnThrIleIysGlnLeuGlnIlyrAlaGlyValIysMetValMet-- 217
      ::::|  ::::|  ::::|  ::::|  ::::|
      1002 ATCTTTCAGAGATTTGACATATACAAACCGGAAAGAAC--ATCAC 1046
      ::::|  ::::|  ::::|  ::::|  ::::|
      218 -----IleGlnAsp-----GlyProGlnAsnThrGlyAlaAspIysProLeuAr 232
      ::::|  ::::|  ::::|  ::::|  ::::|
      1047 TGTGAAGGCGACAGTTGAGGCTGTGCACTGTGAGATGAGATTAAGAAAGCTGCG 1106
      |||||  |||||  |||||  |||||  |||||
      232 gIleThrGlyAspProTyrIysValGlnGlnIlaIysGlnMetValLeuIleuIleAr 252
      ::::|  ::::|  ::::|  ::::|  ::::|
      1107 TGAAGGCTTTGAAGAAATGATATGCTGGCTTTAACACCACTCCGGAATCTTCCAGCCT 1166
      |||||  |||||  |||||  |||||  |||||
      252 gAsp-----GlnGlyGlyPheArgGlnu 260
      ::::|  ::::|  ::::|  ::::|  ::::|
      1167 GTACCCCATCACCAAGTTGGCCGCTTCCGATCATCACTTTATCCAGAGAGAGAT 1226
      ::::|  ::::|  ::::|  ::::|  ::::|
      260 1-----ArgAsnGlnIlyrGly-----SerArgIleGlyGlnGlnI 273
      ::::|  ::::|  ::::|  ::::|  ::::|
      1227 TGTGAATCTCTTATCCCAACCAAGGCTGTGGGCGCATCATGTGGGAAGAGGGCGACA 1286
      |||||  |||||  |||||  |||||  |||||
      273 yIleAspValProIleProArgPheAlaValGlyIleValIleGlyIysAsnIlyGlnu 293
      ::::|  ::::|  ::::|  ::::|  ::::|
      1287 CATCAAAACGCTGGCGAGATTCGCGAGGCTTATCAAGATTCGCCCTGGCGAGGCG-- 1344
      |||||  |||||  |||||  |||||  |||||
      293 tIleIysIysIleGlnAsnAspAlaGlyValArgIleGlnPheIysProAspAspGlyTh 313
      ::::|  ::::|  ::::|  ::::|  ::::|
      1345 ----CCAGACGTACAGGAAAGATGTCATCATCACCGGCGCACCGGAA--GCCAGTT 1397
      |||||  |||||  |||||  |||||  |||||
      313 rThrProAsp-----ArgIleAlaGlnIleThrGlyProProAspArgGlyGlnH 330
      ::::|  ::::|  ::::|  ::::|  ::::|
      1398 CAAGGCCAG----- 1407
      ::::|  ::::|  ::::|  ::::|  ::::|
      330 sAlaIaIaGlnIleIleThrAspLeuLeuArgSerValGlnAlaGlyAsnProGlyr 350
      ::::|  ::::|  ::::|  ::::|  ::::|
      1408 -----GAGCGATCTTTGGGAAA----- 1425
      |||||  |||||  |||||  |||||  |||||
      350 oGlyProGlyGlyrGlyrGlyrGlyrGlyrGlyrGlyrGlyrGlyrGlyrGlyr 370
      ::::|  ::::|  ::::|  ::::|  ::::|
      1426 ----CTGAAGAGGAAACTTTTAAACCCAAAGAAAGTGAAGCTGAAAGCGCATAT 1481
      |||||  |||||  |||||  |||||  |||||
      370 yGlyLeuGlnGlnIlyrPheAsnPheIle----- 378
      ::::|  ::::|  ::::|  ::::|  ::::|
      1482 CAGAGGCGCTTCTCCACAGCTGCGCGGTATTTGGCAAAAGTGGCAACCGTGAACA 1541
      |||||  |||||  |||||  |||||  |||||
      379 ----ValProIlyrGlyrGlyrGlyrGlyrGlyrGlyrGlyrGlyrGlyrGlyr 397
      ::::|  ::::|  ::::|  ::::|  ::::|
      1542 ACTGCAAGAACTTAACCAAGTCAAGTCACTGCTGCTGACCAAAAGCGCAGATGAATA 1601
      ::::|  ::::|  ::::|  ::::|  ::::|
      397 rIleSerGlnIleSerGlyAlaArgIleGlnLeuGlnIlyrSerProProProAsnAlaAs 417

```


[illegible]

```

tissue-specific homolog of poly(pyrimidine tract binding protein.";
RL Mol. Cell. Biol. 20:7463-7479(2000).
CC -I- FUNCTION: Binds to the dendritic targeting element and may play a
CC role in mRNA trafficking (By similarity). Part of a ternary
CC complex that binds to the downstream control sequence (DCS) of the
CC pre-mRNA. Mediates exon inclusion in transcripts that are subject
CC to tissue-specific alternative splicing. May interact with single-
CC stranded DNA from the far-upstream element (FUSE). May activate
CC gene expression.
CC -I- SUBUNIT: Part of a ternary complex containing FUBP2, FUBP1, FUBP2
CC and HNRPL.
CC -I- SUBCELLULAR LOCATION: Nuclear. A small proportion is also found in
CC the cytoplasm of neuronal cell bodies and dendrites (By
CC similarity).
CC -I- TISSUE SPECIFICITY: Detected in neural and non-neural cell lines.
CC -I- SIMILARITY: Contains 4 KH domains.
CC -I- CAUTION: Ref.3 sequence differs from that shown due to numerous
CC framehifts.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isd.ch/announce/).
or send an email to license@isb-sib.ch).
-----
DR EMBL; U94832; AAB53222.1; -.
DR EMBL; AF093747; AAD29861.1; -.
DR EMBL; AF093745; AAD29861.1; JOINED.
DR EMBL; AF093748; AAD29862.1; -.
DR EMBL; U69126; AAC50892.1; ALT_FRAME.
DR Genew; HGNC:6316; KHSRP.
DR MIM; 603445; -.
DR InterPro; IPR004087; KH_dom.
DR InterPro; IPR004088; KH_type_1.
DR Pfam; PF0013; KH; 4.
DR SMART; SMO0322; KH; 4.
DR PROSITE; PS50084; KH_TYPE_1; 4.
KW Transport; mRNA transport; mRNA processing; mRNA splicing;
KW Transcription regulation; Trans-acting factor; Nuclear protein;
KW DNA-binding; RNA-binding; Repeat.
FT DOMAIN 144 208 KH 1.
FT FT 233 299 KH 2.
FT DOMAIN 332 386 KH 3.
FT FT 422 489 KH 4.
FT DOMAIN 569 682 4 X 12 AA IMPERFECT REPEATS.
FT REPEAT 569 580 1.
FT REPEAT 615 626 2.
FT REPEAT 641 652 3.
FT REPEAT 671 682 4.
FT REPEAT 671 682 4.
FT DOMAIN 68 67 GLY/PRO-RICH.
FT DOMAIN 495 549 GLY-RICH.
FT DOMAIN 496 610 ALA/GLY/PRO-RICH.
FT FT 96 96 G->V (IN REF. 2).
FT CONFLICT 96 96 V->G (IN REF. 2).
FT CONFLICT 692 692 G->G (IN REF. 2).
FT CONFLICT 694 694 Q->Q (IN REF. 2).
FT CONFLICT 707 707 Q->A (IN REF. 2).
SQ SEQUENCE 707 AA; 72709 MW; E0758BDDE43BCAB6 CRC64;

Alignment Scores:
Pred. No.: 4.42e-09 Length: 707
Score: 259.00 Matches: 149
Percent Similarity: 33.38% Conservative: 88
Best Local Similarity: 20.99% Mismatches: 284
Query Match: 4.40% Indels: 190
DB: 1 Gaps: 24

US-09-270-437D-8 (1-3283) x FUB2_HUMAN (1-707)
QY 104 CCGCGCCTCACCCGCCAGACACTCTTTGGGCGACAGAAAGCTGCCTCGCGG 163
Db 10 ProProGlyProProProProAlaGlyGlyGlyGlyValAlaGly-GlyValAlaGlyGly 29

```

QY 164 GACAGCTCCTGCTGAAGTCCGCGTACGCTTGTGAGACTACCCGACCGAAGTGGGGCA 223
Db |||
29 yProProProGlyProProGlyAlaGlyAspArgGlyGlyGlyProCysGlyGlyGly 49
QY 224 TCCGGCCCATCGAAGCCCTCTCGGGTAAAGTGAATGCATGGGAAAAATCATGAAGTTG 283
Db yPro----- 50
QY 284 ATTACTAGTCTCTAAAAAGCTAAGAGAGGAAAAATTCAATTGGAACATCCCTCTC 343
Db 51 -----GlyGlyGlySerAlaGlyGlyProSerGly 60
QY 344 ACTCGACGTGGAGGTGTGATGAGCTTTTG-----GCTCAATATGAGACAGTGGAGA 397
Db |||
60 nProProGlyGlyGly- ProGlyIleArgGlyAspAlaPheAlaAspAlaValGlna 80
QY 398 ATGTGGAACAACTCAACACAGACACAGAAACCGCGTTGTCAACGTCAATATGCAACA 457
Db |||
80 rGAlaArgGlnIle----- 84
QY 458 GAGAAAGACAAATAAGCCATGAGAGAGTAAAG-----GGCATCAGTTTGAAACT 511
Db |||
85 -----AlaAlaValIleGlyGlyAspAlaAlaThrThrGlyAsnAsnSerThrProAsp 103
QY 512 ACTCCTTCAGATTTCCTACATCCCGATGAAGAGTGAAGTACCTCCCTTCGCGCCCTCAGC 571
Db |||
103 heGlyPheGlyGlyGlnLysArgGlnLeuGlnuAspGlyAspGlnProGlySerIlySyl 123
QY 572 GAGCCCAAGCTGGGAGCACTCTTCCGCGAGCAAGG-----CACGCCCTGGGGGCA 625
Db |||
123 euAlaSerGlnGlyAspSerIleSerSerGlnLeuGlyProIleIleProProProArg 143
QY 626 CTTCACAGGCGACAGATGATTTCCCGCTGGCGATCTGTGCTCCCAACCCAGTTTGTG 685
Db |||
143 hrSerMetThrGlnIle-----TyrArgValProAspGlyMetValG 157
QY 686 GTGCATCATCGAAGAGAGGCTTGACCATTAAGAACATCATCAAGAGACCGACGCCC 745
Db |||
157 IlyLeuIleIleGlyArgGlyGlyGlnGlnIleAsnIlyIleGlnIleAspSerIlySyl 177
QY 746 GGGTATATTCATAGAAGAAAGAACTCTGAGCTGAGAGAAAGCTGTCAACATCCATG 805
Db |||
177 ySValGlnIle---SerProAspSerGlyGlyLeuProGlnuArgSerValSerLeuThrG 196
QY 806 CCACCCCAAGGGGACTTCTGAGACATCCGCAATGATTTCTTGAATCATGACAGAAA--- 861
Db |||
196 IyAlaProGlnuSerValGlnLysAlaLysMetLeuAspIleValSerArgGly 216
QY 862 -----GAGCAGATGAGACCAAACTAGCCG 886
Db |||
216 rGlyGlyProProGlyGlnPheIleAspAsnAlaAsnGlyGlyGlnAsnGlyThrValG 236
QY 887 AAGAGATTCCTCTGAAAAATTTGGCACACAATGGCTTGTGGAAGACTGATGGAAG 946
Db |||
236 ImlGlnIleMetIleProAla-----GlyLysAlaGlyLeuValIleGlyLysG 252
QY 947 AAGCGAAGATTTGAGAAATTTGAACATGAAACAGGACCAAGATTAACATC-----T 1000
Db |||
252 IyGlyGlnuThrIleLysGlnLeuGlnIleuArgAlaGlyValLysMetIleLeuIleGlna 272
QY 1001 CATCTTTGAGAGATTGAGCATATATACAAACCGGAAAGAACATCAGCTGTAAGGGCAG 1060
Db |||
272 spGlySerGlnAsnThrAsnValAspLysProLeuuhGlyIleIleGlyAspProTyLysV 292
QY 1061 TTGAGGCTGTGCGCACTGTGAGATGAGATTATGAGAAGCTGCGTGAAGCCCTTTGAAA 1120
Db |||
292 aLglnGlnAlaCysGlnMetValMetAspIleLeuArgAsnValThrLysAlaGlyPheG 312
QY 1121 ATGATATGCGGTGTAAACACCCACTCCGAGTATCTTCACGCTGTATCCCAACAC 1180
Db |||
312 IyAsp-----ArgAsnG 316

QY 1181 AGTTTGC-----CCGTCCCGCATATCACTCTTATC 1213
Db |||
316 IuTyGlySerArgIleGlyGlyIleAspValProValProArgHis----- 332
QY 1214 CAGAGCAGAGATTGTGAATCTTTCATCCCAACCCAGAGCTGGGGCCCATCATCGGGA 1273
Db ----- 333
QY 1274 AGAAGGGGACACATCAACAGCTGGAGATTGCGCGAGCTCTTATCAAGATTGCCC 1333
Db |||
340 rGserGlyGlnMetIleLysIleIleGlnAsnAspAlaGlyValArgIleGlnPheLysG 360
QY 1334 CTGGGAAAGGCCACAGACTCAGCGAAGAGATGTCATATATCAACGGGCCACCGAAGCCC 1393
Db |||
360 IAspAspGlyThrGly---ProGlnuIyIleAlaIleAsnIleMetGlyProProAsp---A 378
QY 1394 AGTTCAAGGCCACGAGAGATCTTT----- 1419
Db |||
378 rGyGlnuIleAlaAlaArgIleIleAsnAspLeuLeuGlnSerLeuArgSerGlyProP 398
QY 1420 -----GGGAACTGAAGAGAAACT 1441
Db |||
398 roGlyProProGlyGlyProGlyIleProGlyIleArgGlyArgGlyArgGlyGlnGly 418
QY 1442 TCTTAAACCCCAAGAAAGTGAAGCTGGAAGCGCATATATCAGATGCTCTTCACAG 1501
Db |||
418 snItrGlyProGlyGlyGlnMetThrPheSer-----IleProThrHisLysC 434
QY 1502 CTGCGCGGGTGAATTGCAAGAGTGGCAAGACCGTGAAGCAATGCAAGAACTTAACCAATG 1561
Db |||
434 ySgLyLeuValIleGlyArgGlyGlyGlnuValIySAlaIleAsnGlnIleThrGly 454
QY 1562 CAGAAATCATCGTGCCTGTGACCAAGGCCA-----GATGAATAAGAGAAATGA 1612
Db |||
454 IAspValGlnIleSerArgGlnLeuProProThrGlyIleProThrSerIlySerIleuHei 474
QY 1613 TCGTCAGA-----ATTATCGGCACTTCTTGTACGCAACATGACACAGCGCA 1660
Db |||
474 IelleArgGlySerProGlnGlnIleAspHis-----CysArgGlnLeuIleGlnIle 492
QY 1661 AGATCAGGGGAAAT---TGTAACAAGGTGAAGCAGAGAGAGGAGGAAATACCTCAGGA 1716
Db |||
492 ySIIleGlnuGlyProLeuCysProValGlyProGlyGlyProGlyProAlaGlyP 512
QY 1717 GTGCGCTCAGAGCGCAGCAAGTAGAGCTCCACAGGACCAAGCAAAACGAGATGAATG 1776
Db |||
512 roMetGlyProPheAsnProGlyProPheAsnGlnGlyProProGlyAlaProProHis 532
QY 1777 TAGCCCTTCACACCTGCAGCAATG----- 1802
Db |||
532 IeGlyGlyProProProHisGlnIlyrProProGlnGlyTrGlyAsnThrTyProGlnT 552
QY 1803 -----AGACCAAGCCAGCCAGCCAGATCGGGAGCAACCAAGAGA 1842
Db |||
552 IrgLInProProAlaProHisAspProSerIlySAlaAlaAlaAlaAspProAsn 572
QY 1843 CCATCTGAGGA-----ATGAGAAGCTCGCGAGCGCGCCAGCGAGCTCTCCAGAGC 1893
Db |||
572 IaAlaTrpAlaAlaTyTyTySerHisTyTyTyGlnGlnProProGlyProValPro 590
QY 1894 CTTGAAGACCCCAAGGGGCGAGGAGGGGGGAGAGGTCAAGCCAGGTTTCCGAAGCAC 1953
Db |||
591 -----GlyProAlaProAlaProAlaProAlaProAlaGlnGlyLubPro 606
QY 1954 CGAGCCCGCTCCGCGCCCGCCAG 1977
Db |||
606 roGln- ProProProThrThrGlyGln 613
RESULT 6
ZBP2_CHICK
ID ZBP2_CHICK STANDARD; PRT; 769 AA.
AC Q8UV09;
DT 28-FEB-2003 (Rel. 41, Created)

FT DOMAIN 138 210 KH 2.
 FT DOMAIN 311 380 KH 3.
 FT DOMAIN 394 455 KH 4.
 FT DOMAIN 535 599 KH 5.
 FT VARSPIC 532 532 L-> F (in isoform 2).
 FT VARSPIC 533 606 Missing (in isoform 2).
 FT VARSPIC 533 606 /FTid=VSP_008890.
 FT SEQUENCE 606 AA; 65760 MW; 61f135BB8647C0C CRC64;
 Alignment Scores:
 Pred. No.: 5, 71e-08 Length: 606
 Score: 241.50 Matches: 104
 Percent Similarity: 39.50% Conservative: 86
 Best Local Similarity: 21.62% Mismatches: 178
 Query Match: 4.10% Indels: 113
 Gaps: 16
 DB: 1
 US-09-270-437d-8 (1-3283) x Y475_ARATH (1-606)
 QY 208 GACCAAGAACTGAGCCATCCGCCCATCGAAGCCCTTCGGGTAAAGTGAATTGATGGG 267
 DB 136 AspAspAsnLysGluCysArgLeuValProPheSerGlnSerSerLeuIleGly 155
 QY 268 AAAATCATGAGATTGATTACTCACTCTAAAGCTAAG3--AGCAGAAATTCAG 324
 DB 156 LysAlaGlyGluAsnIleLysArgIleArgArgArgThrArgAlaSerValIleVal 175
 QY 325 ATTCGACATC-----CCTCCATCCTGAGTGGAGGTGTGTGATGACTTTTGCT 378
 DB 176 SerLysAspValSerAspProSerHisValCysAlaMetGluTyrAspAsnValIleVal 195
 QY 379 CAATATGGAGAGTGAAGATGTGAACAAGTCAACACAGACAGAAACCCCTTCG 438
 DB 196 IleSerGlyGluProGluSerValIleGln-----AlaLeuPheAlaValSer 211
 QY 439 AACGTACATATGCAACAGAGAAAGAAATATGACATGAGAGAGTACGCGGAG 498
 DB 212 AlaIleMetTyrLysIleAsnProArgGluAsnIleProLeuAspSerThrSer 229
 QY 499 CAGTTTGAAGACTACTCTTCAAGATTTCCTACATCCCGATGAAGAGTGAAGCTCCCT 558
 DB 230 -----GlnAspValProAlaAlaSerValIleValPro 240
 QY 559 TCGCCCTCCAGCGAGCC-----CAGCGTGGGAGACACTCTCCCGGAG----- 603
 DB 261 LeuGlnGlnGlyAlaGlyValProSerTyrPheAsnAlaLeuSerValSerAspPheGln 280
 QY 607 GGGCAGAGCCCT----- 618
 DB 281 GlyTyrAlaGlnThrAlaAlaAsnProValProValPheAlaSerSerLeuProValThr 300
 QY 619 -----GGGGGCACTTCTCGAGCCAGACAGATGATTCCCGGTGGGATCCCGGTC 669
 DB 301 HisGlyPheGlyGlySerSerArgSerGluGluLeuValPhe-----LysValIleCys 318
 QY 670 CCAACCCAGTTTGTGTGTCATCGAAGAAAGAGGAGCTTGAACATAAGATCATCT 729
 DB 319 ProLeuCysAsnIleMetArgValIleGlyLysGlySerThrIleLysArgIleArg 338
 QY 730 AACGAGACCCAGTCCCGGATGATATTCATGAAAGAAAGAACTCTGAGCTGACAGAG 789
 DB 339 GluAlaSerLysSerCysIleGluValAlaAspSerArgThrLysCysGlyAspAspGlu 358
 QY 790 CCGTGTACCATCATGCCACCCAGAGGAGACTTTCAGACGACGCGCATGATTCCTGAA 849
 DB 359 CysValIleIleValIThrAlaThrGluSerProAspAspMetLysSerMetAlaValGlu 378
 QY 850 -----ATCATGTGAGAAAGAGGAGATGAGACCAACTAGCCGAGAGATTCTCTGAA 903

DB 379 AlaValLeuLeuGlnGlnGlyTyrIleAsnAspGluAspAlaGluAsnValLysMetGln 398
 QY 904 ATCTTGGCACACAAATGCTGTTGGAAGACTGATTGGAAGAAAGAGCAGAAATTTGAG 963
 DB 399 LeuValSerSerLysValIleGlyCysValIleGlyLysSerGlySerValIleAsn 418
 QY 964 AAAATGAGATGAGAAACAGGAGCCAGATACAAATTCGA-----TCTTTCAGAGATTG 1017
 DB 419 GluIleArgLysArgThrAsnAlaAsnIleCysIleSerLysGlyLysAspAspLeu 438
 QY 1018 AGCATATACAAACCCGAAAGAACCATCTACTGTGAAGGACAGATGAGCGCTGGCCAGT 1077
 DB 439 -----ValGluValSerGlyGluValSerSerValArgAsp 450
 QY 1078 GCTGAGATGAGATTATGAAAGAGCTGAGGAGCTTTGAAATGAT----- 1125
 DB 451 AlaLeuIleGlnIleValLeuArgLeuArgGluAspValIleGlyAspLysAspSerVal 470
 QY 1126 -----ATGCTGGCTGTAAACACCCAC 1146
 DB 471 AlaThrArgLysProProAlaArgThrAspAsnLysSerPheLeuSerGlySerSerAsn 490
 QY 1147 TCCGGAATC-----TTCACAGCTGTACCCCATCAGCAG 1182
 DB 491 AlaGlyTyrThrLeuProSerPheMetSerMetAlaSerThrSerGlyPheHisGly 510
 QY 1183 TTTGGCCGCTCCCG-----CATCATCATCTTAT--- 1212
 DB 511 TyrGlySerPheProAlaGlyAspAsnValLeuGlySerThrGlyProTyrSerTyrGly 530
 QY 1213 -----CCAGAGCAGAGATTGTGAATCTTTCATCCACACCCAGCTGTGGCCGCATC 1266
 DB 531 ArgLeuProSerSerSerAlaLeuGluIleLeuIleProAlaHisAlaMetSerLysVal 550
 QY 1267 ATCGGAGAAAGAGGGGACACATCAACAGCTGGAGAGTGGCCGAGCTCTATCAAG 1326
 DB 551 MetGlyLysGlyGlyGlyAsnLeuGluAsnIleArgGlyIleSerGlyAlaMetIleGlu 570
 QY 1327 ATGGCCCTCGGAGAGGCGCACAGCTGACGAGAGATGTCATCATCAACGCGGACCG 1386
 DB 571 IleSer---AlaSerLysThrSerHisGlyAspHisIleAlaLeuLeuSerGlyThrLeu 589
 QY 1387 GAA 1389
 DB 590 Glu 590
 RESULT 8
 PCB3 MOUSE STANDARD; PRT; 339 AA.
 AC P57722;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Poly(I/C)-binding protein 3 (Alpha-CP3).
 GN PCBp3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20396135; PubMed=10936052;
 RA Makeyev A.V., Liehaber S.A.;
 RT Identification of two novel mammalian genes establishes a subfamily
 RT of Kx-domain RNA-binding proteins.";
 RL Genomics 67:301-316(2000).
 CC - FUNCTION: Single-stranded nucleic acid binding protein that binds
 CC preferentially to oligo dC (by similarity).
 CC - SUBCELLULAR LOCATION: Nuclear (by similarity).
 CC - TISSUE SPECIFICITY: Ubiquitous.
 CC - SIMILARITY: Contains 3 KH domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AF176327; AAG09238.1; -
DR HSSP: Q07244; 1KHM.
DR MGD: MGI:1890470; Pcbp3.
DR InterPro: IPR004087; KH_dom.
DR InterPro: IPR004088; KH_type_1.
DR Pfam: PF00013; KH; 3.
DR SMART: SM00322; KH; 3.
DR PROSITE: PSS0084; KH TYPE 1; 3.
KW Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding;
KW Repeat.
FT DOMAIN 13 63 KH 1.
FT DOMAIN 97 150 KH 2.
FT DOMAIN 261 313 KH 3.
SQ SEQUENCE 339 AA; 35958 MW; A6F1C7C176A64F9C CRC64;

Alignment Scores:
Pred. No.: 6.09e-08 Length: 339
Score: 240.00 Matches: 87
Percent Similarity: 41.76% Conservative: 55
Best Local Similarity: 25.59% Mismatches: 105
Query Match: 4.08% Indels: 94
DB: 1 Gaps: 12

US-09-270-437D-8 (1-3283) x PCB3_MOUSE (1-339)

QY 643 ATGATATTCGGCGGATCCGTCGTCGCCACCCAGTTGTGTGTCATCATCGGAAG 702
DB 12 ValThreuthrIleAlaGluLeuMetHisGlyLysGluValGlySerIleIleGlyLys 31
QY 703 GAGGCGCTTGACCATTAAGACATCATAGACAGACCCGCGGAGTAGATCCATCAGA 762
DB 32 LysGlyGluThrValLysLysMetArgGluGluSerGlyAlaArgIleAsnIle----- 49
QY 763 AAAGAGAACTCTGAGCTGCAAGAGAACCTTGCAACATCCATGCCACCCGAGGGGACT 822
DB 50 ---SerGluGlyAsnGlySerProGluArgGlyLeuValThrIleThrGlyProThrAspAlaIle 68
QY 823 TCTGAGAGCAGCGCGATGATCTT-----GAATCATGCAGAAAGAGCCA 867
DB 69 PheLysAlaPheAlaMetIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIle 88
QY 868 GATGAGACCAACTAGCCGAA---GAGATTCCTCTGAAATCTTGAGCAGCAATGGCTTG 924
DB 89 AsnSerProAlaThrSerLysProProValThrLeuArgLeuValAlaProAlaSerGln 108
QY 925 GTTGGAAAGCTGATGAGAAAGAGGCAAAATTTGAAGAAATTTGAACATGAAGAGG 984
DB 109 CysGlySerLeuIleGlyLysGlySerIleLysGluIleArgGluSerThrGly 128
QY 985 ACCAAGATACATCTCATTCTTGACAGATTGAGCATATACACCGGAAAGAACCATC 1044
DB 129 AlaGlnValGlnAlaIleGly-----AspMetLeuProAsnSerThrArgAlaVal 146
QY 1045 ACTGTGAAGGAGCAGTGTGAGGCTGTGCGAGTGTGAGATAGATTTATGAAGAGCG 1104
DB 147 ThrIleSerGlyThrProAspAlaIle-----PheGlnCys 158
QY 1105 CGTGAAGGCTTTGAAAATGATATGCTGTTTAAACCCACTCCGGAT----- 1153
DB 158 sVal-----LysGlnIleCysValValMetLeuGluSerProProLysGlyAlaThr 175
QY 1154 -----ACTTCCAGCCCTGTACCCCGCATCAC----- 1179
DB 175 rIleProLysArgProLysProAlaCysThrProValIlePheAlaGlyGlyGlnAlaTyr 195

QY 1179 ----- 1179
DB 195 rThrIleGlnGlyGlnIleAlaIleProHisProAspGlnLeuThrLysLeuHisGlnLe 215
QY 1180 -----CAGTTTGCCCGTCCCGCAT----- 1200
DB 215 uAlaMetGlnGlnIleThrProPheProProLeuGlyGlnThrAsnProAlaPheProGlyG 235
QY 1201 -----CATCACTTTATCCAGACGAGAGATTGTG----- 1230
DB 235 uLysLeuProLeuHisSerSerGluGluAlaGlnAsnLeuMetGlyGlnSerSerCysLe 255
QY 1231 -----AATCTTTCATCCCAACCCAGGCTGTGGGCGC 1262
DB 255 uAspAlaSerProProAlaSerThrHisGluLeuThrIleProAsnSpleuIleGlyCys 275
QY 1263 CATCATCGGAGAAAGGAGGCGCACATCAACACAGCTGGCGATTCGCGAGGCTCAT 1322
DB 275 sIleIleGlyArgGlnIleGlyThrLysIleAsnGluIleArgGlnMetSerGlyAlaGlnI 295
QY 1323 CAAGATTGCCCCCTCGG---GAGGCGCCAGACGTGCGGAAAGATGTCATCATCACCGG 1379
DB 295 eLysIleAlaAsnAlaThrGluGlySer-----SerGluArgGlnIleThrIleThrG 313
QY 1380 GCCACCGGAGCCCGATTCAGGCGCCGAGGAGCATCTTTGGCAACTGAAGAGGA 1437
DB 313 YThrProAlaAsnIleSerLeuAlaGlnIleLysAlaArgIleuThrSerGln 332

RESULT 9
ID PCB3_HUMAN STANDARD; PRT; 339 AA.
AC P57721;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE POLY(rC)-binding protein 3 (Alpha-CP3).
GN PCB3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20396135; PubMed=10936052;
RA Makeyev A.V., Liebnaber S.A.;
RT "Identification of two novel mammalian genes establishes a subfamily
of Kx-domain RNA-binding proteins";
RL Genomics 67:301-316(2000).
CC -!- FUNCTION: Single-stranded nucleic acid binding protein that binds
CC preferentially to oligo dC (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Contains 3 KH domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AF176329; AAG09240.1; -
DR HSSP: Q07244; 1KHM.
DR Genew: HGNC:8651; PCB3.
DR GO: GO:0005634; C:nucleus; ISS.
DR GO: GO:0003677; F:DNA binding; ISS.
DR GO: GO:0003723; F:RNA binding; NAS.
DR GO: GO:0016071; P:mRNA metabolism; NAS.
DR InterPro: IPR004087; KH_dom.
DR InterPro: IPR004088; KH_type_1.
DR Pfam: PF00013; KH; 3.
DR SMART: SM00322; KH; 3.
DR PROSITE: PSS0084; KH TYPE 1; 3.

RA Dubois E., el Bakkoury M., Glansdorff N., Messenguy F., Pierard A.,
 RA Schreurs B., Vierendeels F.,
 RA Submitted (Aug-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP PARTIAL CHARACTERIZATION.
 RX MEDLINE=99038243; PubMed=9819425;
 RA Mangus D.A., Amrani N., Jacobson A.;
 RT "Pdp1p, a factor interacting with Saccharomyces cerevisiae poly(A)-
 RT binding protein, regulates polyadenylation.";
 RL Mol. Cell. Biol. 18:7383-7396(1998).
 CC -1 SUBUNIT: Interacts with PAB1.
 CC -1 SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1 SIMILARITY: Contains 3 KH domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z36101; CAA85196.1; -.
 CC PIR: S46109; S46109.
 CC Germonline: 138776; -.
 DR SGD: S0000437; PBP2.
 DR InterPro: IPR004087; KH_dom.
 DR InterPro: IPR004088; KH_type_1.
 DR Pfam: PF00013; KH; 3.
 DR SMART: SM00322; KH; 3.
 DR PROSITE: PS50084; KH TYPE 1; 3.
 DR KW Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat.
 FT DOMAIN 66 130 KH 1.
 FT DOMAIN 148 213 KH 2.
 FT DOMAIN 330 394 KH 3.
 SQ SEQUENCE 413 AA; 45782 MW; 638846509BCE1840 CRC64;
 Alignment Scores:
 Pred. No.: 3.06e-07 Length: 413
 Score: 229.50 Matches: 105
 Percent Similarity: 42.20% Conservative: 79
 Best Local Similarity: 24.08% Mismatches: 179
 Query Match: 3.90% Indels: 73
 DB: 1 Gaps: 19
 US-09-270-437d-8 (1-3283) x PBP2_YEAST (1-413)
 QY 523 ATTTCCTACATCCCGATGAGAGGTAGAGCTCCCTTCGCCCCCTCAGGAGCCGCGT 582
 Db 10 IletHrThrThrProHrThrValLeuValSerProAsnThrLeuLysArg--LysLys 28
 QY 583 GGGGACCACTCTTCCCGGAGCA----- 606
 Db 29 GtYgluAspThrSerLgluInleuGluAlaGluIleLysArgValAlaLeuLysAsp 48
 QY 607 -----GGCCAGGCCCTCGGGGCACTTCTCAGGCCAGACAGATT-----GATTCCCG 654
 Db 49 AlaAspSerHisSerAspAsnAspHisAspSerProAspAsnValProSerAspValHis 68
 QY 655 CTGCGGATCGTGTCCCAACCCAGTTGTGTGTCATCATCGAAAGAGGCGCTTGACC 714
 Db 69 LeuArgMetLeuLysLeuValLysHisAlaSerLeuIleValLysLysGlyAlaThr 88
 QY 715 ATTAAGAACCTACTAAGACACCCAGTCCCGGTAGATTCATAGAAAAGAACTCT 774
 Db 89 IleserArgIleLysSerGluThrSerAlaArgIleAsnIleSerAsnValLeuArg-- 107
 QY 775 GGAAGCTGAGAGAGAGCTGTCAACATCCATCCAGCCAGAGGAGACTTGAAGCATGC 834
 Db 108 GlyValProGluArgIleValLysArgGlyThrCysAspAspValAlaLysAlaTyr 127
 QY 835 CGCATGATCTTGAATCATGACAGAAAAGAGGAGATGAGCAAACTAGCCGAA----- 888

Db 128 GlyMetIleValArgAlaLeuLeuGluGluHisGlyLysGluAspAsnGlyLysPile 147
 QY 889 GAGATTCCTCGAAATCTTGACACACATGCTGTGTGGAGACTGATGGAAAAGAA 948
 Db 148 GluIleSerIleAsnLeuLeuIleProHisIleLeuMetGlyCysIleIleGlyLysArg 167
 QY 949 GGCAGAAATTTGAGAAAATTTGAACATGAAACAGGAGCCAGATACATCTCATCTTTG 1008
 Db 168 GlySerArgLeuArgGluIleGluAspLeuSerAlaAlaLysLeuPheAlaSerProAsn 187
 QY 1009 CAGATTTGAGCATATACACCCGAAAGAACCCATCTGGAAGAGGAGCAGATTTGAGCC 1068
 Db 188 Gln-----LeuLeuLeuSerAsnAspArgIleLeuThrIleAsnGlyValProAspAla 205
 QY 1069 TGTCCAGTGTCTGAGATGAGATTAATGAGAGAGCTGCGTGAAGCCCTTGAATGATATG 1128
 Db 206 IletHlleAlaThrPheTyrIleSerGlnThrLeuLeuAsn--PheGluMetGluSer 224
 QY 1129 CTGGCTGTTAACACC--CACTCCGATATCTTC-----TCCAGCCTG 1167
 Db 225 ProGluLysAsnValLysArgSerIleTyrTyrGlnProThrGlnPheAsnSerValLeu 244
 QY 1168 TACCCCATCACAGTTTGGCCCGCTCCGCAT-----CATCACTTATCCAGAG 1218
 Db 245 IleAspHisSerGlnProAsnThrIlePheHisGlnArgAsnHisGlnTyrHisProSer 264
 QY 1219 CAGGAGATTTGTG-----AATCTCTTATCCCAACCCAGGCTGTGGCCGATC 1266
 Db 265 AspyLysLeuLeuSerTyrLysProAsnLysAsnLeuProIleSerSerThr----- 281
 QY 1267 ATCGGAGAAAGAGGGGACACATCAACAGCTGGGG-----AGATTCCCGGAGCTCT 1320
 Db 282 -----LeuLeuSerMetAlaThrProGluThrThrThrAlaSer 294
 QY 1321 ATCAAGATTTGCCCTCGGAGAGGCCGACAG--GTCAAGGAAAGATGATGATCATCAC 1377
 Db 295 ValAlaAsnAlaThrAlaPheGlnProAsnPheValIleProAsnValThrValLeuAsp 314
 QY 1378 GGGCCACCGGAGAGCCAGTTCAAGGCCAGAGGAGCGATCTTTGGAACTGAAAGAGAA 1437
 Db 315 GlyProValIleSerProAlaProGlyAsnHisLeuMetAsnPheValGlnGlnGlu 334
 QY 1438 AACTTCTTTAACCCCAAGAGAGATGAAAGCTGGAAGCCATATCGAGTGCCTCTTCC 1497
 Db 335 IlePheIleAspGluLys----- 340
 QY 1498 ACAGCTGCGGGGTGATGGCAAGGTGGCAAGCCGTGAACGAATGCAACTTAAAC 1557
 Db 341 PheValGlyAsnValIleGlyLysAspGlyLysHisIleAsnSerValLysGluSerThr 360
 QY 1558 AGTGAAGAGTCAATGCTGCTGCTGTGACCAACGCCAGATGAA--AATGAGAAATGATC 1614
 Db 361 GlyCysSerIleIleIle-----GlnAspProValGluIleYserGluAspArg 377
 QY 1615 GTCAAGATTAATGGGCACTTCTTGTGATCCAGACTGCA-----CAGCCGAG 1662
 Db 378 LeuThrIleArgGlyThrPheMetAlaSerGlnAlaIleIleMetLeuIleSerAsnLys 397
 QY 1663 ATCAGGGAATGTACACAGGTGAAGCAGAGCAGAGCAAAATATACCT 1710
 Db 398 Ile--GluIleAspArgSerAsnAlaGluArgLysArgArgSerPro 412
 RESULT 12
 PCBL_HUMAN STANDARD; PRT; 356 AA.
 AC Q15355; Q15157; Q14975;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Poly(UC)-binding protein 1 (Alpha-CP1) (hmrnp-E1) (Nucleic acid
 binding protein SUB2.3).
 GN PCBP1.
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=9531278; PubMed=7607214;
 RA Leffers H., Dejgaard K., Celis J.E.;
 RT "Characterisation of two major cellular poly(ribo)-binding human
 RT proteins, each containing three K-homologous (KH) domains";
 RL Eur. J. Biochem. 230:447-453 (1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=96016208; PubMed=7556077;
 RA Kiledjian M., Wang X., Liebhaber S.A.;
 RT "Identification of two KH domain proteins in the alpha-globin mRNA
 RT stability complex";
 RL EMBO J. 14:4357-4364 (1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymphocytes;
 RX MEDLINE=94203810; PubMed=8152927;
 RA Aasheim H.-C., Loukianova T., Deggerdal A., Smeland E.B.;
 RT "Tissue specific expression and cDNA structure of a human transcript
 RT encoding a nucleic acid binding [oligo(dC)] protein related to the
 RT pre-mRNA binding protein K.";
 RL Nucleic Acids Res. 22:959-964 (1994).
 CC -1- FUNCTION: Single-stranded nucleic acid binding protein that binds
 CC preferentially to oligo dC.
 CC -1- SUBCELLULAR LOCATION: Loosely bound in the nucleus. May shuttle
 CC between the nucleus and the cytoplasm.
 CC -1- TISSUE SPECIFICITY: Abundantly expressed in skeletal muscle,
 CC thymus and peripheral blood leucocytes while a lower expression is
 CC observed in prostate, spleen, testis, ovary, small intestine,
 CC heart, liver, adrenal and thyroid glands.
 CC -1- PM: PHOSPHORYLATED. THE NON-PHOSPHORYLATED FORM(S) EXHIBITED THE
 CC STRONGEST POLY(RC)-BINDING ACTIVITY.
 CC -1- SIMILARITY: Contains 3 KH domains.
 CC -1- CAUTION: There is probably a frameshift error in the nucleotide
 CC sequence of Ref.3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X78137; CAAS5016.1; -;
 DR EMBL; U24223; AAA91317.1; -;
 DR EMBL; Z29505; CAA82631.1; ALT_FRAME.
 DR HSSP; Q07244; 1KHM.
 DR Genew; HGNC:8647; PCBP1.
 DR GK; Q15365; -;
 DR KIM; 601209; -;
 DR GO; GO:0005737; C:cytoplasm; NAS.
 DR GO; GO:0005634; C:nucleus; NAS.
 DR GO; GO:0003723; F:RNA binding; IDA.
 DR GO; GO:0003697; F:single-stranded DNA binding; IDA.
 DR GO; GO:0016071; P:mRNA metabolism; NAS.
 DR InterPro; IPR004087; KH dom.
 DR InterPro; IPR004088; KH_type_1.
 DR Pfam; PF00013; KH; 3.
 DR SMART; SM00322; KH; 3.
 DR PROSITE; PS50084; KH_type_1; 3.
 KW Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding;
 KW Phosphorylation; Repeat.
 FT DOMAIN 13 75 KH 1.
 FT DOMAIN 97 162 KH 2.
 FT DOMAIN 279 343 KH 3.
 FT CONFILICT 205 205 V -> A (IN REF. 2 AND 3).
 SQ SEQUENCE 356 AA; 37526 MW; DC85477576DC5104 CRC64;

Alignment Scores:
 Pred. No.: 5.74e-07 Length: 356
 Score: 225.00 Matches: 82
 Percent Similarity: 40.40% Conservative: 61
 Best Local Similarity: 23.16% Mismatches: 123
 Query Match: 3.82% Indels: 88
 DB: 1 Gaps: 11
 US-09-270-437D-8 (1-3283) x PCBP1_HUMAN (1-356)
 QY 619 GGGGCACTTCTCAGGCGCAGACGATTTCCCGCTGGGAGTCTGTCGCCACCCAG 678
 Db 4 GlyValThrGluSerGlyLeuAsnValThrLeuThrIleArgLeuMetHisGlyLeu 23
 QY 679 TTTGTGGTGCATCATCGGAAGAAGGAGCTTGACCAATAAGACATCATCAAGACAG 738
 Db 24 GluValGlySerIleIleGlyLeuValGlySerValValValArgGluGluSer 43
 QY 739 CAGTCCGGGTAGATATTCATAGAAAAGAACTCTGAGCTGCAGAGACCTGTCAAC 798
 Db 44 GlyAlaArgIleAsnIle-----SerGluGlyAsnCysProGluArgIleIleThr 60
 QY 799 ATTCATGCGACCCGAGAGGGGACTTCTGAGCATGCGGCATGATTTGAAATCATGAG 858
 Db 61 LeuThrGlyProThrAsnAlaIlePheLeuAlaPheAlaMetIleLeuAspLeuVal 80
 QY 859 AAGAGGCGAGTAGAGCCAACTA-----GCCGAAGAGATTCTCTG 900
 Db 81 GluAspIleAsnSerMetThrAsnSerThrAlaIleAsnArgProProValThrLeu 100
 QY 901 AAAATCTGGCACAACATGCTTGTGGAAAGCTGATTTGAAAAGAGCAGAAATTG 960
 Db 101 ArgLeuValValProAlaThrGlnCysGlySerLeuIleGlyGlyCysLeuIle 120
 QY 961 AAGAAATTGAACATGAAGAGGAGCAGATTAACATTCATCTTGACAGATTGAGC 1020
 Db 121 LysGluIleArgGluSerThrGlyAlaGlnValGlnValAlaGly-----AspMetLeu 138
 QY 1021 ATATACACCCGGAAGAACCATCATCTGTAGAGGACAGTTGAGGCTGTGCAGTGTCT 1080
 Db 139 ProAsnSerThrGluArgAlaIleThrIleAlaGlyValProGlnSerValThrGluCys 158
 QY 1081 GAGATAGACATT-----ATGAAGAAGTGGCTGAGGCTTTGAAAATGATATG 1128
 Db 159 ValLysGlnIleCysLeuValMetLeuGluThrLeuSerGlnSerProGlnGlyArgVal 178
 QY 1129 CTGGCTGTTAACACCCACTCCGAGTACTTCCAGCTG----- 1167
 Db 179 MetThrIleProGlyGlnProMetProAlaSerSerProValIleCysAlaGlyGln 198
 QY 1168 -----TACCCCAT----- 1176
 Db 199 AspArgCysSerAspAlaValGlyTyrProIleAlaThrHisAspLeuGluGlyProPro 218
 QY 1177 -----CACCACTTGGCCCG----- 1191
 Db 219 LeuAspAlaTyrSerIleGlnGlyGlnHisThrIleSerProLeuAspLeuAlaTyrLeu 238
 QY 1192 -----TTCCCGCATCATGAC----- 1206
 Db 239 AsnGlnValAlaArgGlnGlnSerHisAspAlaMetMetHisGlyGlyThrGlyPheAla 258
 QY 1207 -----TCTTATCCGAG----- 1218
 Db 259 GlyIleAspSerSerSerProGluValGlyGlyTyrThrAlaSerLeuAspAlaSerThr 278
 QY 1219 CAGGAATTTGATATCTTATCCCAACCCAGGCTGTGGGCCCATCATCGGAGAGAG 1278
 Db 279 GlnThrThrHisGluLeuThrIleProAsnAlaLeuIleGlyCysIleIleGlyArgGln 298
 QY 1279 GGGGACACATCAACAGCTGGGAGATTCGCGGAGGCTCTTCAAGATTGCC---CCT 1335
 Db 299 GlyAlaAsnIleAsnGluIleArgGlnMetSerGlyAlaGlnIleLeuIleAlaAsnPro 318

```

QY 1336 GCGGAGGCCCGACAGCTCAGCGGAGAAAGATGTCATCATCACCAGGCCACCGGAGGCCAG 1395
Db 319 ValGluGlySer-----SerGlyArgGlnValThrIleThrGlySerAlaIleSerIle 336
QY 1396 TTCAGAGCCCGAGGAGCGATCTTTGGAGAACTGAAGAGAGAA 1437
Db 337 SerIeuAlaGlnIlyrLeuIleAlaSerAlaArgIeuSerSerGlu 350

RESULT 13
PCBI_RABIT STANDARD; PRT; 356 AA.
ID PCBI_RABIT STANDARD; PRT; 356 AA.
AC 019048;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Poly(rC)-binding protein 1 (Alpha-CPI) (hnrnp-B1).
GN PCBP1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99207122; PubMed=10101190;
RA Thiele B.O., Berger M., Hutch A., Reimann I., Schwarz K., Thiele H.;
RT "Tissue-specific translational regulation of alternative rabbit
RT 15-lipoxygenase mRNAs differing in their 3'-untranslated regions.";
RL Nucleic Acids Res. 27:1828-1836 (1999).
CC -1- FUNCTION: Single-stranded nucleic acid binding protein that binds
CC preferentially to oligo dC (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- SIMILARITY: Contains 3 KH domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ003023; CA05814.1; -.
DR HSPF: Q07244; 1KHV.
DR InterPro: IPR004087; KH dom.
DR InterPro: IPR004088; KH_type_1.
DR Pfam: PF00013; KH; 3.
DR SMART: SM00322; KH; 3.
DR PROSITE: PS00084; KH_type_1; 3.
DR Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding; Repeat.
KW DOMAIN
FT DOMAIN 13
FT DOMAIN 97
FT DOMAIN 279
FT DOMAIN 343
SQ SEQUENCE 356 AA; 37497 MW; 6D1A261276CA206D CRC64;

Alignment Scores:
Pred. No.: 5.74e-07 Length: 356
Score: 225.00 Matches: 82
Percent Similarity: 40.40% Conservative: 61
Best Local Similarity: 23.16% Mismatches: 123
Query Match: 3.82% Indels: 88
DB: 1 Gaps: 11

US-09-270-437D-8 (1-3283) x PCBI_RABIT (1-356)
QY 619 GGGGACCTTCTCAGCCAGACAGATTGATTTCCGTCGGATCCTGTGTCCTCCACCCAG 678
Db 4 GlyValThrGluSerGlyLeuAsnValThrLeuThrIleArgIeuSerMetHsGlyTys 23
QY 679 TTTGTTGGTCATCATCGAAGAGAGGGCTTGAACATTAAGAACTCATTAAGCAGACC 738
Db 24 GluValGlySerIleIleGlyLysGlyGluSerValIleArgGluGluSer 43

```

```

QY 739 CAGTCCCGGGTATATTCATGAAAAGAACTGTGAGCTGCAGAGACCTGTACC 798
Db 44 GlyAlaThrGlyLeuSerIle-----SerGluGlyAsnGlyProGluArgIleIleThr 60
QY 799 ATCCATGCCACCCCGAGAGGAGCTTTCAGACATCCCGCATGATCTTGAATCATGAGC 858
Db 61 LeuThrGlyProThrAsnAlaIlePheLysAlaPheAlaMetIleIleAspLysLeuGlu 80
QY 859 AAAGAGCGAGTGAACCAACTA-----GCCGAAGAGATTCCTGTG 900
Db 81 GluAspIleAsnSerSerMetThrAsnSerThrAlaIleAspArgProProValThrLeu 100
QY 901 AAAATCTTGCAACAATGCTGCTGGTGAAGATGATGAAAAGAGAGAAATTTG 960
Db 101 ArgIeuValAlaProAlaThrGlnCysGlySerIleGlyLysGlyCysLysIle 120
QY 961 AAGAAATATTAACATGAACAGAGGAGCCAAAGTAAACATTCATCTTTCAGATTTGAGC 1020
Db 121 LysGluIleArgGluSerThrGlyAlaGlnValGlnValIleGly-----AspMetIeu 138
QY 1021 ATATTAACCCGGAAGAACCATCATCTGTGAAGGACACAGTTGAGGCTGTGCCAGTGT 1080
Db 139 ProAsnSerThrGluArgAlaIleThrIleAlaGlyValProGlnSerValThrGluCys 158
QY 1081 GAGATGAGATT-----ATGAAGAAGCTGCGGTGAGGCTTGAAAATGATATG 1128
Db 159 ValLysGlnIleCysLeuValMetLeuGluThrLeuSerGlnSerProGlnIleArgVal 178
QY 1129 CTGGCTGTAAACCCACCATCCGAGATCTTCGCCAGCTG 1167
Db 179 MetThrIleProTyrGlnProMetProAlaSerSerProValIleCysAlaGlyGln 198
QY 1168 -----TACCCCAT----- 1176
Db 199 AspArgCysSerAspAlaAlaGlyTyrProHisAlaThrHisAspLeuGluGlyProPro 218
QY 1177 -----CACCATTTGGCCCG----- 1191
Db 219 LeuAspAlaTyrSerIleGlnIleGlnIleThrIleSerProLeuAspLeuAlaLysLeu 238
QY 1192 -----TTCCGCATCATC----- 1206
Db 239 AsnGlnValAlaArgGlnGlnSerHisPheAlaMetMetHisGlyGlyThrGlyPheAla 258
QY 1207 -----TCTATCCAG----- 1218
Db 259 GlyIleAspSerSerSerProGluValLysGlyTyrTrpAlaSerIeuAspAlaSerThr 278
QY 1219 CAGAGATTGTGAATCTTTCATCCACACCGCTGTGGGCCCATCTCGGAGAGAG 1278
Db 279 GlnThrThrHisGlnMetThrThrIleProAsnAsnLeuIleGlyCysIleIleGlyArgGln 298
QY 1279 GGGGACATGATGAACAGCTGGGAGATTCGCGGAGCTCTTCAAGATTTGCC--CCT 1335
Db 299 GlyAlaAsnIleAsnGlnIleArgGlnMetSerGlyAlaGlnIleLysIleAlaAsnPro 318
QY 1336 GCGGAGGCCCGACAGCTCAGCGGAGAAAGATGTCATCATCACCAGGCCACCGGAGGCCAG 1395
Db 319 ValGluGlySer-----SerGlyArgGlnValThrIleThrGlySerAlaIleSerIle 336
QY 1396 TTCAGAGCCCGAGGAGCGATCTTTGGAGAACTGAAGAGAGAA 1437
Db 337 SerIeuAlaGlnIlyrLeuIleAlaSerAlaArgIeuSerSerGlu 350

RESULT 14
VGIN_HUMAN STANDARD; PRT; 1268 AA.
ID VGIN_HUMAN STANDARD; PRT; 1268 AA.
AC Q00341; Q9UCY3;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vgillin (High density lipoprotein-binding protein) (HDL-binding
DE protein).

```

ON HDLBP OR HBP OR VGL.
 GS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92291094; PubMed=1318310;
 RA McKnight G.L., Reasner J., Gilbert T., Sundquist K.O., Hokland B.,
 RA McKernan P.A., Champagne J., Johnson C.J., Bailey M.C., Holly R.,
 RA O'Hara P.J., Oram J.F.;
 RT Cloning and expression of a cellular high density lipoprotein-binding
 RT protein that is up-regulated by cholesterol loading of cells.;
 RT J. Biol. Chem. 267:12131-12141(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RX MEDLINE=22388257; PubMed=1247932;
 RA Strassburg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
 RA Altshul S.F., Zeeberg B., Buecaw K.H., Scheffer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldi M.F., Casavari T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loughellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunatrine P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Mizny D.C., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=96184515; PubMed=8605996;
 RA Kugler S., Gruneweller A., Probst C., Klinger M., Muller P.K.,
 RA Kruse C.;
 RT "Vaglin contains a functional nuclear localisation sequence and is
 RT present in both the cytoplasm and the nucleus.";
 RT FEBS Lett. 382:330-334(1996).
 CC -1- FUNCTION: APPEARS TO PLAY A ROLE IN CELL STEROID METABOLISM.
 CC IT MAY FUNCTION TO PROTECT CELLS FROM OVER-ACCUMULATION OF
 CC CHOLESTEROL.
 CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
 CC -1- SIMILARITY: Contains 14 KH domains
 CC -1- SIMILARITY: STRONG. TO C.ELEGANS C08H9.2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/1sb-sib.ch>).
 CC or send an email to license@1sb-sib.ch.
 CC -----
 DR EMBL; M64098; AAA5962.1; -;
 DR EMBL; BC001179; AAH01179.1; -;
 DR PIR; A44125; A44125.
 DR PDB; 1VTG; 03-APR-96.
 DR PDB; 1VTH; 03-APR-96.
 DR Genew; HGNC:4857; HDLBP.
 DR MIM; 142695; -;
 DR GO; GO:0005886; C:plasma membrane; TAS.
 DR GO; GO:0008289; F:lipid binding; TAS.
 DR GO; GO:0008203; P:cholesterol metabolism; TAS.
 DR InterPro; IPR004087; KH_dom.
 DR InterPro; IPR004088; KH_type_1.

[illegible]

```

QY 961 AAGAAATTTGAACATGAAACAGGAGCCAGATTAACATTCATCTTTGACGATTGAGC 1020
Db 605 LysylserineArginylglutamineThrylle-----AspLeuPro 618
QY 1021 ATATACCAACCCG--GAAAGAACATCAGTGTGAAGGC-----ACAGTTGAGGCC 1068
Db 619 AlaGluamserAsnserGluThrIleIleIleThrGlyLysArgAlaAsnCysGluAla 638
QY 1069 TGTGCCAGTGTGAGATAGATATTATGAAGAGCTGCGTGAAGGCTTTGAAAATGATATG 1128
Db 639 AlaArgSerArgIleLeuSerIleGlnLysAspLeuAlaAsnIleAlaGlu----- 655
QY 1129 CTGGCTGTAAACACCCACTCCGGAATCTTCCTCCAGCCTGTACCCCATCAGCAGTTGAC 1188
Db 655 ----- 655
QY 1189 CCGTTCGCCGATCATCTTATTCAGAGCAGAGATTGATCTGTTCATCCCAAC 1248
Db 656 -----ValGluValSerIleProAla 662
QY 1249 CAGCGTGTGGCCGATCATCGGGAAGAGGCGACACATCTAAACAGCTGCGCAGATTTC 1308
Db 663 LysLeuHisAsnSerLeuIleGlyThryGlyArgLeuIleArgSerIleMetGluGlu 682
QY 1309 GCCGGAGGCTGTATCAAGATTGCGCCCTGGGGAAGGCCAGACGTACGAAAGATGCTC 1368
Db 683 CysGlyGlyValHisIleHisPheProValGluGlySer-----GlySerAspThrVal 700
QY 1369 ATCATCAGCGGCGCAGCGGAGCCAGTTCAGAGCCGAGGAGCGAGATCTTTGGAAACTG 1428
Db 701 ValIleArgGlyProSerSerAspValGluLysAlaLysGluGlnLeu-----HisLeu 719
QY 1429 AAGAGGAAACTCTTTTAAACCCAAAGAGAGTGAAGTGAAGCGATTCAGAGTGG 1488
Db 720 AlaGluGlu-----LysGlnThrLysSerPheThrValAspIleArgAla 734
QY 1489 CCCTCTCCACAGCTGCGCGGAGTGGCAAGGTCGCAACCGCTGACGAACTGACG 1548
Db 735 LysProGluThryHisLysPheLeuIleGlyLysGlyGlyLysIleArgLysValArg 754
QY 1549 AACTTAACCAAGTGAAGATGATCTGCTCTGACCAAGCCAGATGAAATGAGGAA 1608
Db 755 AspSerThrGlyAlaArgValIlePheProAlaAlaGluAspLysAspGln----- 772
QY 1609 GTGATCGTCAAGATTATCGGCGACTTCTTGTGACGCAACAGTCCAGAGCGAAAGATCAG 1668
Db 773 ---LeuIleThrIleIleGlyLysGluAspAlaValArgGluAlaGlnLysGluLeuGln 791
QY 1669 GAAATGTGTACACAGGTG 1686
Db 792 AlaLeuIleGlnAsnLeu 797

```

```

EN [2]
RP FUNCTION.
RX MEDLINE=22120219; PubMed=12124753;
RA Kumar D.V., Nigam A., St John P.A.;
RT "Role of Nova-1 in regulating alpha2N, a novel glycine receptor
RL splice variant, in developing spinal cord neurons.";
RU J. Neurobiol. 52:156-165(2002).
CC -1- FUNCTION: Acts to regulate alternative splicing in neurons by
CC binding pre-mRNA in a sequence-specific manner to activate exon
CC inclusion. It binds specifically to the sequence UCAUY (by
CC similarity). Acts to regulate a novel glycine receptor alpha-2
CC chain splice variant (alpha-2N) in developing spinal cord.
CC -1- SIMILARITY: Contains 3 KH domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, AY262017; AAP20872.1; -.
CC InterPro; IPR004087; KH_dom.
CC InterPro; IPR004088; KH_type_1.
CC Pfam; PF00013; KH; 3.
CC SMART; SM00322; KH; 3.
CC PROSITE; PS50084; KH_TYPE_1; 3.
CC Antigen; Nuclear protein; RNA-binding; Repeat.
CC KW Antigen; Nuclear protein; RNA-binding; Repeat.
CC FT DOMAIN 26
CC FT DOMAIN 48
CC FT DOMAIN 146
CC FT DOMAIN 248
CC FT DOMAIN 326
CC FT NON_TER 474
CC SQ SEQUENCE 474 AA; 48443 MW; 70D0926B5BEC131 CRC64;

Alignment Scores:
Pldg. No.: 1.41e-06 Length: 474
Score: 219.50 Matches: 105
Percent Similarity: 35.77% Conservative: 66
Best Local Similarity: 21.97% Mismatches: 190
Query Match: 3.73% Indels: 117
Gaps: 16

US-09-270-437d-8 (1-3283) x NOAL_RAT (1-474)
QY 562 CCCCTCAGAGCCAGCGTGGGACCACTCTCCCGGAGCAAGCCAGCCCTGGG 621
Db 21 ProProAspSerArgLysArgProLeuGlnAlaProProGluAlaGly--SerThrLys 39
QY 622 GGCACCTTCAGGCCAGAGATGATTTCCCGTGGGATCTGCTCCCACTGTT 681
Db 40 ArgThrAsnThrGlyLysGlyLysGlyLysPheLeuValLeuIleProSerThrAla 59
QY 682 GTTGTCGATCATCGGAAAGAGGCTTGACCATTAAGACATCACTAAGCAGCCAG 741
Db 60 AlaGlySerIleIleGlyLysGlyGlyGlnThrIleValGlnLeuGlnLysGlnThrGly 79
QY 742 TCCCGGTAGATTCATTAAGAAAGAGAC-----TTGGAGCTGCAAGAGCCCTGTC 795
Db 80 AlaThrIleLysLeuSerLysSerLysAspPheThrProGlyThrGluArgValCys 99
QY 796 ACATCATCCAGCCAGCCAGGAGGAGCTTCGAAGCATGCGCATGATTTGAAATCAT 855
Db 100 LeuIleGlnGlyThrIleGluAlaLeuAsnAlaValHisGlyPheIleAlaGluLysIle 119
QY 856 CAGAAAGAGGCAATGAGACCAACTAGCCGAAAGATTCCTCTG----- 900
Db 120 ArgGluMetProGlnAsnValAlaLysThrGluProValSerIleLeuGlnProGlnThr 139

```

Search completed: July 23, 2004, 11:06:57
Job time : 58.8809 secs

```
QY 901 -----AAATCTTGCGACAAATGGCTTGTTGGA 930
Db 140 ThrValAsnProAspArgIleLysGlnValLysIleIleValProAsnSerThrIaGly 159
QY 931 AGACTGATTTGAAAAGAGCGCAAAATTGAAAGAAATTGAAACATGAAACAGGACCAAG 990
Db 160 LeuIleIleGlyLysGlyIaIaThrValLysIaIleMeGlnGlnSerGlyIaIaTrp 179
QY 991 ATAAACATCTCATCTTTGACAGATTGACATATACACCCGGAAGAAGAACATCATCTGTG 1050
Db 180 ValGlnLeuSerGlnLysProAsp--GlyIleAsnLeuGlnGlnArgValIaIaThrVal 198
QY 1051 AAGGCGACAGTTGAGGCGCTGTGCCAGTGTGAGATAGATATGATGAAAGAGCTGGCTGAG 1110
Db 199 SerGlyGlnProGlnGlnAsnArgLysIaIaValGlnLeuIleIleGlnLysIleGln 218
QY 1111 GCCTTTGAAAATGATATGCTGCTGTAAACACCACTCCGGAATACTTCTCCAGCCTGTAC 1170
Db 219 -----AspProGlnSerGlySerCysLeuAsnLysSer 229
QY 1171 CCCCATCACAGTTTGCGCGGTTCCCGCATCATCAC-----TCTTATCCAGACGAGGAG 1224
Db 230 TyrIaAsnValIaThrGlyProValIaIaAsnSerAsnProThrGlyPheProTyrIaAsn 249
QY 1225 ATTGTGAATCTCTTCATCCCAACCCAGGCTGTGGGCGCC--ATCATCGGAGAAAGGGG 1281
Db 250 ThrIaGlnVal--LeuProThrIaIaIaIaIaIaGlyLeuLeuGlyHis----- 266
QY 1282 GCACACATCAAAACAGCTGGCGAGATTG----- 1308
Db 267 AlaAsnLeuIaGlyValIaIaIaIaPheProIaIaValLeuSerGlyPheThrGlyAsnAsp 286
QY 1308 ----- 1308
Db 287 LeuValIaIaIeThrSerIaIaLeuAsnThrIeuaIaSerTyrglyTyraAsnLeuAsnThr 306
QY 1309 -----GCCGAGCCTCTATCAAGATTGCCCTGGCG 1338
Db 307 LeuGlyLeuGlyLeuSerGlnIaIaIaIaIaIaThrGlyIaIeuaIaIaIaIaIaIaSer 326
QY 1339 GAAGGCCCAAGCGTCAGCGAAAGAGTGTATCATCAACCGGCGCACCGGAAGCCCAAGTTC 1398
Db 327 AlaAsnProIaIaIaIaIaIaIaIaAsnLeuIaIaThrTyraIaSerGlnIaIa----- 344
QY 1399 AAGGCCAGGAGCGATCTTTGGG----- 1422
Db 345 SerIaSerGlySerThrIaGlyGlyThrIaGlyThrPheIaLeuGlySerLeuIaIa 364
QY 1423 -----AAACTGAAAGAGAAACTTCTTAACCCCAAGAAAGAGTGAAGCTGGAAGCG 1476
Db 365 AlaIaIaThrIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 384
QY 1477 -----CATATCAGAGTGCC 1491
Db 385 IleLeuGlyThrGlnLysSerThrAspGlySerLysAspValValGlnIleIaIaValPro 404
QY 1492 TCTTCCACAGCTGGCGGCGGATTTGGCAAGGTGGCAAGACCGTGAACGAATGCGACAGAC 1551
Db 405 GlnAsnLeuValIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 424
QY 1552 TTAAACAGTGCAGAGATCATGCTGCTCGT-----GACCAAGCGCAAGTGAAGAAATGAG 1605
Db 425 LeuThrGlyIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 444
QY 1606 GAAGTATGCTGCAAAATTATCGGCACTTCTTGAACCACTGACAGCACTGACAGCGCAAGATC 1665
Db 445 ArgLys--ValThrIleThrGlyThrProIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 463
QY 1666 AAGGAAATTTGTAACAACAGTGAAGCAGGAGACAGAAATACCTCAGGAGATC 1719
Db 464 -----ThrGlnArgIleThrTyrglnGlnGlyVal 473
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 23, 2004, 10:55:00 ; Search time 156.961 Seconds

(without alignments)
13198.757 Million cell updates/sec

Title: US-09-270-437D-8

Perfect score: 5886

Sequence: 1 ggcagcgagagcgagcgagga.....aaccttgaaatgttattc 3283

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ n2p.model -DEV=x1p
-O=/cpg2_1/USPROO_pool_P/US9270437/runat_23072004_095739_27425/app_query.fasta_1.5582
-DB=SPRMBL_25 -QFMT=fstanc -SUFFIX=fstpc -MINMATCH=0.1 -DOFCL=0 -DOEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US9270437 @CGN 1.1 287 @runat 23072004 095739 27425 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEJOB -NEG_SCORES=0 -NAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -MARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPRMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	2838	48.2	556 4 Q9Y6M1	Q9Y6M1 homo sapien

2	2123	36.1	545	11	Q7TOP9	Q7TGF9 mus musculus
3	1965.5	33.4	594	13	Q73932	Q73932 xenopus lae
4	1965	33.4	593	13	Q57526	Q57526 xenopus lae
5	1951.5	33.2	576	11	Q42254	Q42254 gallus gall
6	1938	32.9	577	11	Q8CGX0	Q8CGX0 rattus norv
7	1936	32.9	577	11	Q88477	Q88477 mus musculus
8	1935	32.9	577	11	Q9NZ18	Q9NZ18 homo sapien
9	1929	32.8	577	11	Q8BRH1	Q8BRH1 mus musculus
10	1929	32.8	577	11	Q8U0S9	Q8U0S9 mus musculus
11	1890.5	32.1	582	13	Q9PW80	Q9PW80 brachydiano
12	1875	31.9	579	11	Q8C2J9	Q8C2J9 mus musculus
13	1875	31.9	579	11	Q9CPN8	Q9CPN8 mus musculus
14	1868	31.7	579	4	Q00425	Q00425 homo sapien
15	863	14.7	580	5	Q81GK4	Q81GK4 drosophila
16	858.5	14.6	573	5	Q81R99	Q81R99 drosophila
17	856	14.5	566	5	Q9VZ69	Q9VZ69 drosophila
18	478.5	8.1	828	5	Q21605	Q21605 caenorhabdi
19	337	5.7	100	4	Q86V81	Q86V81 homo sapien
20	275.5	4.7	641	5	Q9BLA0	Q9BLA0 caenorhabdi
21	274	4.7	589	5	Q17935	Q17935 caenorhabdi
22	274	4.7	611	5	Q17936	Q17936 caenorhabdi
23	272	4.6	568	10	Q9LXF5	Q9LXF5 arabidopsis
24	263.5	4.4	557	5	Q9C553	Q9C553 arabidopsis
25	261.5	4.4	557	5	Q23487	Q23487 caenorhabdi
26	260.5	4.4	169	11	Q7TP50	Q7TP50 rattus norv
27	255	4.3	313	4	Q96EP6	Q96EP6 homo sapien
28	254.5	4.3	510	5	Q7Z145	Q7Z145 caenorhabdi
29	254	4.3	680	5	P91393	P91393 caenorhabdi
30	250.5	4.3	653	13	Q7ZXS1	Q7ZXS1 xenopus lae
31	250	4.2	398	3	Q74919	Q74919 schizosacch
32	250	4.2	762	10	Q81NT9	Q81NT9 oryza sativ
33	246.5	4.2	833	10	Q9FMF0	Q9FMF0 arabidopsis
34	246	4.2	351	4	Q8N9K6	Q8N9K6 homo sapien
35	244.5	4.2	1281	13	Q7ZTN2	Q7ZTN2 xenopus lae
36	241	4.1	610	13	Q8AX84	Q8AX84 xenopus lae
37	240	4.1	313	10	Q81Z82	Q81Z82 arabidopsis
38	237.5	4.0	313	13	Q7ZYK5	Q7ZYK5 brachydiano
39	237.5	4.0	370	11	Q8BSB0	Q8BSB0 mus musculus
40	237.5	4.0	542	10	Q857G1	Q857G1 oryza sativ
41	237	4.0	371	11	Q85C44	Q85C44 mus musculus
42	235	4.0	644	10	Q81DV1	Q81DV1 arabidopsis
43	235	4.0	644	10	Q9FNK3	Q9FNK3 arabidopsis
44	234.5	4.0	479	10	Q9X171	Q9X171 arabidopsis
45	233.5	4.0	1208	13	Q7ZU44	Q7ZU44 brachydiano

ALIGNMENTS

RESULT 1

Q9Y6M1 PRELIMINARY; PRT; 556 AA.
ID Q9Y6M1
AC Q9Y6M1;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hepatocellular carcinoma autoantigen.
OS Pez.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99207072; PubMed=10190901;
RA Zhang J.Y., Chan E.K., Peng X.X., Tan E.M.;
RT "A novel cytoplasmic protein with RNA-binding motifs is an autoantigen
in human hepatocellular carcinoma.";
RT J. Exp. Med. 189:1101-1110(1999).
CC -1- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR EMBL; AF057352; AAD31596.1; -
DR GO; GO:0005737; Cytoplasm; TAS.
DR GO; GO:0003723; rRNA binding; TAS.
DR GO; GO:0007345; P.embryogenesis and morphogenesis; TAS.

DR GO:0009386; P:translational attenuation; TAS.
DR InterPro: IPR004087; KH dom.
DR InterPro: IPR004088; KH_type_1.
DR InterPro: IPR00504; RNA_rec_mot.
DR Pfam: PF00013; KH; 4.
DR Pfam: PF00016; KH; 2.
DR SMART: SMO0322; KH; 4.
DR SMART: SMO0360; RM; 2.
DR PROSITE: PS50084; KH_type_1; 4.
DR PROSITE: PS50102; RM; 2.
DR SEQUENCE 556 AA; 61842 MW; 18DEFB10043DCC4 CRC64;
Alignment Scores:
Pred. No.: 4.72e-200 Length: 556
Score: 2838.00 Matches: 556
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 48.22% Indels: 0
Gaps: 0
US-09-270-437D-8 (1-3283) x Q9Y6M1 (1-556)
QY 70 ATGATGAACAAGCTTTATCATCGGGAACCTGAGCCCGCGCTCACCGCGAGCACTCGG 129
Db 1 MetMetAsnLysLeuTyrIleGlyAsnLeuSerProAlaValThrAlaAspAspLeuArg 20
QY 130 CAGCTCTTTGGGAGACAGAAAGCTGCTCCCTGGCGGACAGCTCTGCTGAAGTCCGGCTAC 189
Db 21 GlnLeuPheGlyAspArgLysLeuProLeuAlaGlyGlnValLeuLeuLysSerGlyTyr 40
QY 190 GCCTTCGAGACTACCCGACGAGAACTGGGCACTCCGCGCATCGAGACCTCTCGGAT 249
Db 41 AlaPheValAspTyrProAspGlnAsnTyrAlaIleArgAlaIleGlnThrLeuSerGly 60
QY 250 AAAGTGAATTCATGAGGAATCATGGAATGATTAATCTCAGTCTTAAAAAGCTTAAG 309
Db 61 LysValGlnLeuHisGlyLysIleMetGluValAspTyrSerValSerLysLysLeuArg 80
QY 310 AGCAGAAATTCAGATTGAAACATCCTCTCTCAGCTGCACTGGAGAGTGTGGATGGA 369
Db 81 SerArgLysIleGlnIleArgAsnIleProProHisLeuGlnTyrGlnValLeuAspGly 100
QY 370 CTTTGGGCTCATATGCGGACAGTGAGATGTGGAACAAGTCAACACAGACAGAAACC 429
Db 101 LeuLeuAlaGlnTyrGlyThrValGlnAsnValGlnGlnValAsnThrAspThrGlnThr 120
QY 430 GCCGTGTCAAGTCATATGCAATATGCAACAGAGAAAGCAAAAATAGCCATGAGAAAGCTA 489
Db 121 AlaValValAsnValThrTyrAlaThrArgGlnGlnValLysIleAlaMetGlnLysLeu 140
QY 490 AGCGGGCATCAGTTGAGAACTATCTCTTCAGATTCTCTCATCCCGAGTGAAGAGTG 549
Db 141 SerGlnHisGlnPheGlnLysAsnTyrSerPheLysIleSerTyrIleProAspGlnGlnVal 160
QY 550 AGTCCCTTCGCGCCCTCAGAGGAGCCGCGTGGGAGCAACTCTCCCGAGAGCAAGGC 609
Db 161 SerSerProSerProGlnArgAlaGlnArgGlyAspHisSerSerArgGlnGlnGly 180
QY 610 CAGCGCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCGCTGCGGATCTTGTC 669
Db 181 HisAlaProGlyGlyThrSerGlnAlaArgGlnIleAspPheProLeuArgIleLeuVal 200
QY 670 CCCACCCAGTTGTGTGTCATCATCGGAAGAGAGGGCTTGACCTAAAGAACATCACT 729
Db 201 ProThrGlnPheValGlyAlaIleIleGlyLysGlnGlyLeuThrIleLysAsnIleThr 220
QY 730 AAGCAGAACCCAGTCCCGGAGATATCCATGAGAAAGAGAACTCTGAGCTGAGAGAAAG 789
Db 221 LysGlnThrGlnSerArgValAspIleHisArgLysGlnLysSerGlyAlaIleGlnLys 240
QY 790 CCGTGCACCATCCATGCCAACCCAGAGGGGACTTGTGAAGCATGCGGCATGATTTTGAA 849
Db 241 ProValThrIleHisAlaThrProGlnGlyThrSerGlnLysCysArgMetIleLeuGln 260

QY 850 ATCATGCAAAAGAGCAGATGAGACCAACTAGCCGAGAGATTCCTTGAAATCTTG 909
Db 261 IleMetGlnLysGlnAlaAspGlnThrLysLeuAlaGlnLysIleProLeuLysIleLeu 280
QY 910 GCACACAAATGGCTGTGTTGGAAGCTGATTTGGAAAAGAGCGCAATTTGAAAGAAAT 969
Db 281 AlaHisAsnGlyLeuValGlyArgLeuIleGlyLysGlnGlyArgAsnLeuLysLysIle 300
QY 970 GAACATGAAACAGGAGCCAAATATCAATCTCATCTTTCAGATTGAGCATATACAC 1029
Db 301 GlnHisGlnThrGlyThrLysIleThrIleSerSerLeuGlnAspLeuSerIleTyrAsn 320
QY 1030 CCGGAAAGAACCATCATCTGTGAGAGGACAGTTGAGGCTGTGCTGAGATGAG 1089
Db 321 ProGlnThrThrIleThrValLysGlyThrValGlnLysAlaSerAlaGlnIleGln 340
QY 1090 ATTATGAAGAGCTGCGTGGAGGCTTTGAAAATGATATGCTGAGTGAACACCACTCC 1149
Db 341 IleMetLysLysLeuArgGlnAlaPheGlnAsnAspMetLeuAlaValAsnThrHisSer 360
QY 1150 GGATTAATCTTCAGGCTGTACCCCATCAGCAGTTGGCCGTTCCGCAATCATCTCT 1209
Db 361 GlyTyrPheSerSerLeuTyrProHisHisGlnPheGlyProPheProHisHisSer 380
QY 1210 TATCCAGAGCAGAGATTGTGAATCTTTCATCCCAACCGCTGTGGGCCATCATC 1269
Db 381 TyrProGlnGlnGlnIleValAsnLeuPheIleProThrGlnAlaValAlaIleIle 400
QY 1270 GGGAAAGAGGGGACACATCAAAAGCTGCGAGATTGCGCGAGCTCATCAAGATT 1329
Db 401 GlyLysLysGlyAlaHisIleLysGlnLeuAlaArgPheAlaGlyAlaSerIleLysIle 420
QY 1330 GCCCTGGGAGAGGCCCAAGCCTGACGAAAGATGTGCATCATCATCAGCGGCCACCGAA 1389
Db 421 AlaProAlaGlnGlyProAspValSerGlnArgMetValIleIleThrGlyProProGln 440
QY 1390 GCCCAGTTCAAGGCCAGGAGCGATCTTTGGAAAATGAAAGAGAAATCTTTTAAC 1449
Db 441 AlaGlnPheLysAlaGlnGlyArgIlePheGlyLysLeuLysGlnGlnAsnPheAsn 460
QY 1450 CCCAAAGAAAGATGAAGCTGGAAGCGCATATCAAGAGCCCTTCCACAGCTGGCCGG 1509
Db 461 ProLysGlnGlnValLysLeuGlnAlaHisIleArgValProSerSerThrAlaGlyArg 480
QY 1510 GTGATTTGGAAAGTGGGAGACCGTGAACGAATGTCAAGAACTTAACAGTCCAGAGTTC 1569
Db 481 ValIleGlyLysGlyGlyLysThrValAsnGlnLeuGlnAsnLeuThrSerAlaGlnVal 500
QY 1570 ATCGTCCCTGTGACCAACGCGCAGATGAATAAGAAAGTATGCTCAGAAATTATCGGG 1629
Db 501 IleValProArgAspGlnThrProAspGlnAsnGlnGlnValIleValArgIleIleGly 520
QY 1630 CACTTCTTTGCTAGCCAGACTCAGACGCGCAAGATCAGGGAATTTACAACAGTGAAG 1689
Db 521 HisPhePheAlaSerGlnThrAlaGlnArgLysIleArgGlnIleValGlnGlnValLys 540
QY 1690 CAGCAGAGCAGAAATACCTCAGAGGAGTGGCTTCACAGCGAGCAAG 1737
Db 541 GlnGlnGlnGlnLysTyrProGlnGlyValAlaSerGlnArgSerLys 556
RESULT 2
Q7TOR9 PRELIMINARY; PRT; 545 AA.
AC Q7TOR9; (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner U., Shemmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carrinci P., Prange C.,
RA Rata S.S., Loquellano N.A., Peters G.J., Adamson R.D., Mullaly S.J.,
RA Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Wotley K.C., Hale S., Garcia A.M., Gay L.J., Halys S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley K.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC054552; AAH54552.1; -.
KW Hypothetical protein.
SQ
SEQUENCE 545 AA; 59679 MW; 88DD35FF30DF3091 CRC64;

Alignment Scores:
Pred. No.: 1.92e-147 Length: 545
Score: 2123.00 Matches: 418
Percent Similarity: 87.76% Conservative: 12
Best Local Similarity: 85.31% Mismatch: 10
Query Match: 36.07% Indels: 50
DB: 11 Gaps: 2

US-09-270-437D-8 (1-3283) x Q7TQF9 (1-545)
QY 73 ATGAACAAGCTTATACATCGGGAACCTGAGCCCGCGCTCACCGCGGACGACCTCCGCGAG 132
Db 1 MetAsnIysLeuYrIleGlyAsnLeuSerProAlaValThrAlaAspAspLeuAArgLn 20
QY 133 CTCTTTGGGAGACGAGAGCTGCCCTGCGGAGACAGGTCTGCTGAAGTCGGGCTAGCGC 192
Db 21 LeuPheGlyAspArgIysLeuProLeuAlaGlyGlnValLeuLeuYssSerGlyTyrAla 40
QY 193 TTGCTGAGACTAATCCCGACCGAAGCTGGGCGATCCGGGCGCATGAGACCTCTCGGGTAA 252
Db 41 PheValAspTyrProAspGlnAsnTyrAlaIleArgAlaIleGlyThrLeuSerGlyLys 60
QY 253 GTGGAATTCGATGGGAAAATCATGGAAGTTGATTACAGTCTCTCTCTCTCTCTCTCTCTCT 312
Db 61 ValGluLeuHisGlyIysIleMetGluValAspTyrSerValSerIysIysLeuAArgSer 80
QY 313 AGGAAATTCAGATTGGAACATCCCTCTCAGCTCAGTGGAGAGGTGTGTGATGACTT 372
Db 81 ArgArgIleGlnIleArgAsnIleProProHisLeuGlnIleThrIleValLeuAspGlyLeu 100
QY 373 TTGGCTCATATATGGGACAGTGGAGAAATGGGAACAAGTCAACACAGACACAGAAACCGCC 432
Db 101 LeuAlaGluTyrGlyThrValGluAsnValGluGlnValAsnThrAspThrGluThrAla 120
QY 433 GTTGTCAAGCTCACATATGCAACAGAGAAGAAAGAAAATATAGCCATGAGAAAGTAAAGC 492
Db 121 ValValAsnValThrIleMetThrArgGlnGlnAlaIysLeuAlaIleGluIysLeuSer 140
QY 493 GGGCATCACTTTGAGAACTACTCTTCAAGATTTCTTACATCCCGGATGAAGAGTGAGC 552

```

```

Db 141 GlyHisGlnPheGluAspTyrSerPheYsIleSerTyrIleProAspGluGlnValSer 160
QY 553 TCCCTCTGGCCCTTCAGCGAGCCGACGGTGGGACCACTTCCCGGAGCAAGGCCAC 612
Db 161 SerProSerProProHisArgAla-----ArgGluGlnGlyHis 173
QY 613 GCCCTGGGGGACATTTCTCAGGCCAGACAGATTGATTTCCCGCTGGCGGATCTGGTCCC 672
Db 174 GlyProGlySerSerSerGlnAlaArgGlnIleAspPheProLeuAArgIleLeuValPro 193
QY 673 ACCGAGTTTGTGGTCATCATCGAAGAAAGGAGGAGCTTGACCAATTAAGACATCATGAG 732
Db 194 ThrGlnPheValGlyAlaIleIleGlyLysGluLeuThrIleYsAsnIleThrIys 213
QY 733 CAGACCAAGTCCCGGATGATATCATTAAGAAAGAAAGAACTGTGAGCTGCAGAGAGCCT 792
Db 214 GlnThrGlnSerArgValAspIleHisArgLysGlnSerSerGlyAlaIleGluIysPro 233
QY 793 GTCAACCATTCATGCCACCCCGAGAGGAGCTTCTAAGCATGCGCATGATTTCTGAATC 852
Db 234 ValThrIleHisAlaThrProGluGlyThrSerGlnAlaCysArgMetIleLeuGluIle 253
QY 853 ATGCAGAAAGAGCGAGATGAGACCAACTGCGGAGAGATTCCTCTGAATAATCTTGACA 912
Db 254 MetGlnYsGlnAlaAspGlnThrIleYsLeuAlaGluGlnValProLeuYsIleLeuAla 273
QY 913 CACAAATGCTGTGGTGAAGACTGATTTGAAGAAAGAGAGCAAAATTTGAAGAAATTTGAA 972
Db 274 HisAsnGlyPheValGlyArgLeuIleGlyLysGluIleYsGlnLeuYsIleGluIle 293
QY 973 CATGAACAGAGAACCAAGATTAACATCTCATCTTTGAGAGATTTGAGCATATACACCCG 1032
Db 294 HisGluThrGlyThrIleYsIleThrIleSerSerLeuGlnAspLeuSerIleTyrAsnPro 313
QY 1033 GAAAGAACCATCACTGCTGAAGGAGCAAGTGAAGGAGCCGTGCGCATGCTGATGATGAGATT 1092
Db 314 GluArgThrIleThrValAlaGlyGlyThrIleGlnAlaCysAlaAsnAlaGluIleGluIle 333
QY 1093 ATGAAGAGCTGCGTGAAGGCTTTGAAATGATATGCTGCTGCTGTAAC----- 1140
Db 334 MetIleYsLeuArgGlnAlaPheGluAsnAspMetLeuAlaValAsnGlnAlaAsn 353
QY 1140 ----- 1140
Db 354 LeuIleProGlyLeuAsnLeuSerAlaLeuGlyIlePheSerThrGlyLeuSerValLeu 373
QY 1141 -----ACC 1143
Db 374 ProProProAlaGlyProArgGlyValProProSerProProTyrHisProPheAlaThr 393
QY 1144 CACTCGGATTACTTCCAGCCTGTACCCCATCACAGTTTGCCCGGTCCTCCGCGATCAT 1203
Db 394 HisSerGlyTyrPheSerSerLeuYrProHisHisHisPheGlyProPheProHisHis 413
QY 1204 CACTCTTATCCAGAGACGAGATTTGAAATCTTTATCCCAACCAAGGCTGTGGCGCC 1263
Db 414 HisSerTyrProGluGlnGlnThrValSerLeuPheIleProThrGlnAlaValGlyAla 433
QY 1264 ATCATGGGAGAAAGAGGGGAGACACATCAAAACGCTGGCGAGATTCGCCGAGGCTCATAC 1323
Db 434 IleIleGlyYsLeuYsGlyAlaHisIleYsGlnLeuAlaArgPheAlaGlyAlaSerIle 453
QY 1324 AAGATTGCCCTGCGGAGAGCCGACAGTCAAGGAAAGAGTGTCTCATCACCGGAGCCA 1383
Db 454 LysIleAlaProAlaGluGlyProAspValSerGlnArgMetValIleIleThrGlyPro 473
QY 1384 CCGGAAGCCCAAGTTCAAGGCCCAAGGAGCG 1413
Db 474 ProGluAlaGlnPheYsValSerAlaYs 483
RESULT 3
073932
ID 073932 PRELIMINARY; PRT; 594 AA.

```

AC 073932; (TREMblrel. 07, Created)
 DT 01-AUG-1998 (TREMblrel. 07, last sequence update)
 DT 01-OCT-2003 (TREMblrel. 25, last annotation update)
 DE vgl RNA binding protein variant D.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 OC Xenopodinae; Xenopus.
 NC NCB1_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Havin L., Git A., Elisha Z., Oberman F., Yaniv K.,
 RA Pressman Schwartz S., Standart N.M., Yisraeli J.K.;
 RL Genes Dev. 0:0-0(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9828351; PubMed=9560341;
 RA Deshler J.O., Hight M.I., Adrason T., Schnapp B.J.;
 RT "A highly conserved RNA-binding protein for cytoplasmic mRNA
 localization in vertebrates."
 RL Curr. Biol. 8:489-496(1998).
 CC -1- SIMILARITY: CONTAINS 4 KH DOMAINS.
 DR EMBL; AF064634; AAC18598.1; -;
 DR EMBL; AF055923; AAC41285.1; -;
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR InterPro; IPR004087; KH_dom.
 DR InterPro; IPR004088; KH_type_1.
 DR InterPro; IPR005054; RNA_rec_mot.
 DR Pfam; PF00013; KH; 4.
 DR Pfam; PF00076; rrm; 2.
 DR SMART; SM00322; KH; 4.
 DR SMART; SM00360; RRM; 2.
 DR PROSITE; PSS0084; KH_type_1; 4.
 DR PROSITE; PSS0102; RRM; 2.
 SQ SEQUENCE 594 AA; 65643 MW; 5ACEAYBF0856D6 CRC64;

Alignment Scores:
 Pred. No.: 7,68e-136 Length: 594
 Score: 1965.50 Matches: 388
 Percent Similarity: 78.50% Conservative: 72
 Best Local Similarity: 66.21% Mismatches: 91
 Query Match: 33.39% Indels: 35
 Db: 13 Gaps: 5

US-09-270-437D-8 (1-3283) x 073932 (1-594)

QY 73 ATGAACAAGCTTTAATCATCGGAACTGAGCCCGCGCGCGAGCACTCCGCGAG 132
 Db 1 MetAshlyleuTyrlleGlyshnleuSerGluAenValSerProThraspleuGluSer 20
 QY 133 CTCCTTTGGGACAGAGAGCTGCCCTGGCGGACAGAGCTCTGTGAAGTCCGCTACGCC 192
 Db 21 LeuPhelyGluSerlyleProPhethrGlyGlnPhelVallysserGlyTyrla 40
 QY 193 TTTCGTGACTAOCOCGACGAGAACTGGGCACTCGGCCATCGAGACCTCTCGGTAA 252
 Db 41 PheValAspCyseroAspGluThrPalaMetLysAlaIleAspThrleuSerGlyLys 60
 QY 253 GTGCAATTCATGGAAGAAATCATGAAAGTTGATTACTCACTCTTAAAAAGCTAAGAGC 312
 Db 61 ValGluLeuHlsGlyLysValIleGluValGluHlsSerValProLysAlaGlnAAsSer 80
 QY 313 AGCAAAATTCAGATTGGAATCATCTCTCTCACCTGACGTGGAGAGGTGTGGATGACTT 372
 Db 81 ArgLysLeuGlnIleArgAsnIleProProHlsIleuGlnIlePrlValLeuAspSerLeu 100
 QY 373 TTGGCTCAATATGGGACAGTGGAGATGTGGACAAAGTCAACAGACAGAAACCGCC 432
 Db 101 LeuAlaGlnTyrllyThrValGluAsnCyserGluGlnValAsnThrgluSerGluThrAla 120
 QY 433 GTTGCAAGCTCACATATGCAACAAGAGAAAGAAATATGCAATGAGGAAAGTAAAGC 492

Db 121 ValValAsnValThrTyrlaAsnlyGluHlsAlaArgGlnGlyLeuGluLysLeuAsn 140
 QY 493 GGCATCATGTTTGAAGACTACTCTCAAGATTCTTACATCCCGATGAGAGATGAGC 552
 Db 141 GlyTyrlleuGluAsnTyrlSerLeuLysValThrTyrlleProAspGluMetAlaThr 160
 QY 553 TCCCTTCGCCCCCTCAGCGAGCCGAGCGGAGCACTTCCCGGAGAGCGCCAC 612
 Db 161 ProGlnAlaProserGlnIleuGlnIleProGlnIleGlnIleHlsPProGlnIlyArg 180
 QY 613 -----GCCCTGGGGGCGACTTCTCAGGCC 636
 Db 181 ArgLysPheGlylnArglyProAlaArgGlnGlySerProGlyAlaAlaAlaArgPro 200
 QY 637 AGA---CAGATTGATTTCCCGCTGCGAGTCTGTCCACCCAGTTTGTGGTGCATC 693
 Db 201 LysProGlnThrgluValProLeuHlyrMetLeuValProThrglnPheValGlyAlaTle 220
 QY 694 ATCGAAAGAGGCGTTGACCATTAAGACATCTCAATAGCAGACCCAGTCCCGGTAGAT 753
 Db 221 IleGlyLysGluGlyAlaThrIleArgAsnIleThrgLysGlnThrglnSerLysLeasp 240
 QY 754 ATCCATGAAAGAGAACTCGGAGCTGACAGAAAGCTGTACACATCCATGCCACCCCA 813
 Db 241 IleHlsArgLysGluAsnAlaGlyAlaAlaGlnLysProIleThrlleHlsSerThrPro 260
 QY 814 GAGGGACATTCGAAAGCATCGCGCATGATCTTGAATCATGACAGAAAGGACAGATGAG 873
 Db 261 GluGlyCyserAlaAlaCyserIleleMetGlnIleMetGlnIlyGluIleGlnAsp 280
 QY 874 ACCAAATAGCCGAAAGAAATTCCTTGAATCTTGGACACAAATAGCGTTGTGGAGGA 933
 Db 281 ThrLysPheThrgluGlnIleProLeuLysIleLeuAlaHlsAsnAspPheValGlyArg 300
 QY 934 CTGATTGAAAGAGAGGAGAAATTTGAAGAAATTTGAACATGAACAGGAGCAAGATA 993
 Db 301 LeuIleGlyLysGluGlyArgAsnLeuLysIleGluGlnAspThraspThrlystle 320
 QY 1054 GGCACAGTTGAGCGCTGTGCCAGTGTGATAGAGATTATTAAGAGCTGGTGAGGCC 1113
 Db 341 GlySerIleGluProCyserAlaLysAlaGluGlnIleMetLysLysIleArgGluSer 360
 QY 1114 TTTGAAATGATATAGCTGGCTGTAAACACCACTCC----- 1149
 Db 361 TyrlGluAsnAspIleAlaAlaMetAsnLeuGlnAlaHlsLeuIleProGlyLeuAsnLeu 380
 QY 1150 -----GATACCTTC-----TCAGCTGTGACCCCATGACAGATTGGC--- 1188
 Db 381 AsnAlaLeuGlyLeuPheProserSerSerGlyMetProProProserValGlyVal 400
 QY 1189 -----CCGTCCCGCATCATCACTCTTAATCCAGAGCAGAGATT 1227
 Db 401 ProserProThrSerSerThrseryrProProPheGlyGlnIleProGluSerGluThr 420
 QY 1228 GTGAATCTTTCATCCCAACCCAGAGCTGTGGCGCATATCGGAGAGAGGGGACAC 1287
 Db 421 ValHlsLeuPheIleProAlaLeuAlaValGlyAlaIlelellyLysGlnIlyGlnHls 440
 QY 1288 ATCAAAACAGCTGCGAGATTCGCGGAGCTCTATCAAGATTGCGCTCGGAGAGCCCA 1347
 Db 441 IlelyGlnLeuSerArgPheAlaGlyAlaSerIlelyIleAlaProAlaGluGlyPro 460
 QY 1348 GACGTACGGAAGAGATGTCATCATCACCGGGCCACCGGAAGCCAGATTCAAGGCCAG 1407
 Db 461 AspAlaLysLeuArgMetValIleIleThrglyProProGluAlaGlnPheLysAlaGln 480
 QY 1408 GAGCGATCTTTGGGAACTGAAAGAGAAACCTCTTAAACCCAAAGAGAGTAAAG 1467
 Db 481 GlyArgIleTyrllyLysLeuLysGluGlnAsnPhePheGlyProLysGluGlnValLys 500

QY 1468 CTGGAAGCCATATAGAGTCCCTCTTCCACAGCTGGCCGGGATTTGGCAAGTGGC 1527
 Db |||||
 Db 501 LeuGluThrHisIleLeuValProSerTyrIleAlaGlyArgValIleGlyIleGly 520
 QY 1528 AAGACCGTGAAGCACTGTGACACTTAAACAGTGCAGAAAGTATCGTGGCTGTGACCA 1587
 Db |||||
 Db 521 LysThrValAsnGluLeuGlnAsnLeuThrSerIleAlaGluValValProArgAspGln 540
 QY 1588 ACGCAGATGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1647
 Db |||||
 Db 541 ThrProAspGluAsnAspGluValValValValValValValValValValValValVal 560
 QY 1648 ACTGCACAGCCGCAAGATCAGGAAATTTGTACACAGTGAAGCAGCAGCAGCAGCAGC 1707
 Db |||||
 Db 561 LeuAlaGlnArgGlyIleGlnGlnIleLeuAlaGlnValAlaArgArgGlnGlnGlnGln 580
 QY 1708 CCTCAGGAGTCCGCTCA 1725
 Db |||||
 Db 581 GlnLysThrValGlnSer 586

RESULT 4

ID 057526 PRELIMINARY; PRT; 593 AA.
 AC 057526;

DT 01-JUN-1998 (Tremblere, 06, Created)
 DT 01-JUN-1998 (Tremblere, 06, Last sequence update)
 DT 01-OCT-2003 (Tremblere, 25, Last annotation update)
 DE KH domain-containing transcription factor B3.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 OC Xenopodidae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92249652; PubMed=1577195;
 RA Pfaff S.L., Taylor W.L.;
 RT "Characterization of a Xenopus oocyte factor that binds to a
 developmentally regulated cis-element in the TFIID gene";
 RL Dev. Biol. 151:306-316(1992).
 RU [2]
 RP SEQUENCE FROM N.A.
 RA Griffin D., Taylor W.L.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]

RP SEQUENCE FROM N.A.
 RA Havin L., Git A., Elisha Z., Oberman F., Yaniv K.,
 RA Pressman Schwartz S., Standart N.M., Yisraeli J.K.;
 RL Genes Dev. 0:0-0(1998).
 CC -1- SIMILARITY: CONTAINS 4 KH DOMAINS.
 DR EMBL: AF042353; AAB97457.1; -
 DR EMBL: AF042353; AAC18597.1; -
 DR GO: 0003676; F: nucleic acid binding; IEA.
 DR InterPro: IPR004087; KH_dom.
 DR InterPro: IPR004087; KH_type_1.
 DR InterPro: IPR005054; RNA_rec_moc.
 DR Pfam: PR00013; KH; 4.
 DR Pfam: PR00076; rtm; 2.
 DR SMART: SM00322; KH; 4.
 DR SMART: SM00360; RRM; 2.
 DR PROSITE: PS00804; KH_TYPE_1; 4.
 DR PROSITE: PS0102; RRM_2.
 DR PROSITE: PS00030; RRM_RNP_1; FALSE NEG.
 SQ SEQUENCE 593 AA; 65385 MW; 5A5AB4B4A1D5D5F7 CRC64;

Alignment Scores:

Pred. No.: 8.35e-136
 Score: 1965.00
 Percent Similarity: 79.27%
 Best Local Similarity: 67.01%
 Query Match: 33.38%
 Db: 13

Length: 593
 Matches: 388
 Conservative: 71
 Mismatches: 84
 Indels: 36
 Gaps: 6

US-09-270-437d-8 (1-3283) x 057526 (1-593)
 QY 73 ATGAACAAGCTTTTACATGGAGCACTGAGCCCGCTACCGCCGACGATCCGGCAG 132
 Db |||||
 Db 1 MetAsnLysLeuTyrIleGlyAsnLeuSerGlnValSerProProAspLeuGlnSer 20
 QY 133 CTCTTTGGGGACAGAGAGCTCCCTGGCGGAGCAGTCTGCTGAGTCCGGCTACGCG 192
 Db |||||
 Db 21 LeuPheLysGlnSerLysIleProPheThrGlyGlnPheLeuValLysSerGlyTyrIle 40
 QY 193 TTGCGGACCTACCCCGACCAAGATCGGCGCATCCGCGCATTCGAGACCTCTCGGGTAAA 252
 Db |||||
 Db 41 PheValAspCysProAspGluThrThrPheLeuValAlaIleAspThrLeuSerGlyLys 60
 QY 253 GTGGAATTCAGTGGGAAATATCATGAAATGTTATTAATCACTGCTCTAAAGCAATGAGAG 312
 Db |||||
 Db 61 ValGluLeuHisGlyLysValIleGluValGlnHisSerValProLysArgGlnArgSer 80
 QY 313 AGGAAATTCAGATTCGAAACATCCCTCTCAGCTGAGTGGAGGTGGATGACTT 372
 Db |||||
 Db 81 ArgLysLeuGlnIleArgAsnIleProProHisLeuGlnIleProGluValLeuAspSerLeu 100
 QY 373 TTGCTCAATATGGGACAGTGGAGATGTGGAATGTGGAATCAACAGACAGAAACGCC 432
 Db |||||
 Db 101 LeuAlaGlnTyrGlyThrValGluAsnCysGlnValAlaSerThrAspSerGluThrAla 120
 QY 433 GTTGTCAAGCTCATATGCAACAGCAAGAAAGCAAAATAGCATGAGAGTAAAGC 492
 Db |||||
 Db 121 ValValAsnValThrTyrAlaAsnLysGlnHisAlaArgGlnGlyLeuGlnLysLeuAsn 140
 QY 493 GGGCATGATGTTGGAACATCTCTTCAAGATTTCTTACATCCGAGTGAAGAGTGAAGC 552
 Db |||||
 Db 141 GlyTyrGlnLeuGlnValSerLeuValValThrTyrIleProAspGluMetAlaThr 160
 QY 553 TCCCTCTGGCCCTCAGCGACGCCAGCGT----- 582
 Db |||||
 Db 161 ProGlnSerProSerGlnGlnLeuGlnGlnProGlnGlnHisProGlnGlyArgArg 180
 QY 583 -----GGGAGCACCTCTCCCGGAGCAGAGCCAGCCCTGGGGGCACTTCCACAGCC 636
 Db |||||
 Db 181 GlyPheGlyGlnArgGlyProAlaArgGlnGly---SerProGlyAlaAlaAlaArgPro 199
 QY 637 AGA---CAGATTGATTTCCCGCTGCGGATCTGCTGCCCAACGATTGTTGGTCCATC 693
 Db |||||
 Db 200 LysProGlnSerGlnValProLeuArgMetLeuValProThrGlnPheValGlyAlaIle 219
 QY 694 ATCGAAAGAGGGGCTGACCCATTAAGAAAGATCACTTAAGACGCCAGTCCGGGTATGAT 753
 Db |||||
 Db 220 IleGlyLysGlnGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAsp 239
 QY 754 ATCCATAGAAAGAGAACTCTGAGCTGCAGAGAGCCGTGCATCCATCCATCCACCCCA 813
 Db |||||
 Db 240 IleHisArgLysGlnAsnAlaGlyAlaAlaGluLysProIleThrIleHisSerThrPro 259
 QY 814 GAGGGGACTTCTGAGCATGCGCATGATCTTGAATCATGACAGAAAGCAGCATGAG 873
 Db |||||
 Db 260 GlnGlyCysSerAlaIaIaCysLysIleIleMetGlnIleMetGlnLysGlnAlaGlnAsp 279
 QY 874 ACCAACTAGCCGAAAGATTCCTCTGAAATTTTGGACACACATGGCTTGGTGAAGA 933
 Db |||||
 Db 280 ThrLysPheThrGlnIleIleProLeuLysIleLeuAlaHisAsnAsnPheValGlyArg 299
 QY 934 CTGATTGGAAGAGAGCAGAAATTTGAAGAATGAAATGAAATGAAATGAAATGAAATGAA 993
 Db |||||
 Db 300 LeuIleGlyLysGlnLysArgAsnLeuLysIleGlnGlnAspThrAspThrLysIle 319
 QY 994 ACAATTCATCTTTCAGGATTTGACATATACACCCGAGAAAGAACATCACTGTGAAG 1053
 Db |||||
 Db 320 ThrIleSerProLeuGlnAspLeuThrLeuTyrAsnProGluArgThrIleThrValLys 339
 QY 1054 GGCACAGTTGAAGGCTGTGCCAGTCTGAGATAGAGATTTAAGAAAGCTGGGTGAGGCC 1113
 Db |||||

```

Db      340 GlySerIleGluThrGlySalAlaGluGluValMetIleGlySerIleArgGluSer 359
QY      1114 TTGAAAATGATATGCTGCTGTTTAAACCACTCC----- 1149
      360 TyGluAsnAspIleAlaIleMetAsnLeuGlnIleHisIleProGlyLeuAsnLeu 379
QY      1150 -----GGATCTTC-----TCCAGCTGTACCCCATCAACAGTTTGGC--- 1188
      380 AsnAlaLeuGlyLeuPheProProSerSerSerIleMetProProProSerAlaGlyVal 399
Db      1189 -----CCGTTCCCGCATCATCACTTTATCCAGACGAGAGATT 1227
QY      400 SerSerProThrThrThrSerAlaSerTyrProProPheGlyGlnProGluSerGluThr 419
Db      1228 GTGATCTCTTCATCCCAACCAAGCTGTGGGCGCCCATCATCGGAGAGAGAGGCGGCAC 1287
QY      420 ValHisIleuPheIleProAlaIleuAlaValAlaIleIleGlyIleGlyIleGlnHis 439
      1288 ATCAAAACGCTGGAGATTCGCCGAGACCTCTATCAAGATTGCCCTGCGGAAAGCCCA 1347
QY      440 IleuGlnIleuSerGluPheAlaGlyAlaSerIleIleAlaProAlaGluGlyPro 459
Db      1348 GACGTCACGGAAGAGATGTCATCATCAACCGGCGCACCGGAAAGCCAGTTCAAGCCAG 1407
QY      460 AspAlaIleuSerGluMetValIleIleThrGlyProProGluAlaGlnPheIleSalGln 479
Db      1408 GGACGGATCTTTGGGAAACTGAAAGAGAAACTCTTTAACCCCAAGAAAGAACTGAAG 1467
QY      480 GlyArgIleTyrGlyIleuIleuSerGlnIleuAsnIlePheGlyProIleGluIleVal 499
Db      1468 CTGGAAGCGCATATCAGAGTGCCCTCTTCCACAGCTGGCCGGGTGATTGGCAAGSTGGC 1527
QY      500 LeuGlnAlaHisIleuValProSerTyrAlaAlaGlyArgValIleGlyIleGlyIle 519
Db      1528 AAGACCGTGAAGCACTGACAGAACTTAAACAGTGCAGAGATCATGTGCTGTGACCAA 1587
QY      520 LysThrValAsnGluLeuGlnAsnIleuThrSerAlaGlnValIleValProArgAspGln 539
Db      1588 AGCCGAGTGAAGAAATGAGAAATGATCTGTCAGAAATTAATTCGGGCACTTCTTCTGCGCAG 1647
QY      540 ThrProAspGluAsnAspGlnValValIleValIleIleThrGlyHisPheTyrAlaSerGln 559
      1648 ACTGCACGCGCAGAGATCAGCAAAATTTGTAACAAGTGAACAGAGCAGAGCAGAA 1704
QY      560 LeuAlaGlnArgIleIleGlnIleIleGlnIleAlaGlnValArgArgGlnGlnGln 578
Db
RESULT 5
042254 PRELIMINARY; PRT; 576 AA.
AC 042254;
DT 01-JAN-1998 (TRMBLrel. 05, Created)
DT 01-JAN-1998 (TRMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TRMBLrel. 25, Last annotation update)
DE Zipcode-binding protein.
GN ZBP1.
OS Gallus gallus (Chicken).
OC Archaeoptera; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97220007; PubMed=9121465;
RA Ross A.F., Oleynikov Y.S., Kislaukis E.H., Taneja K.L., Singer R.H.;
RT "Characterization of a beta-actin mRNA zipcode-binding protein.";
RL Mol. Cell. Biol. 17:2158-2165(1997).
CC -1- SIMILARITY: CONTAINS 4 KH DOMAINS.
EMBL AF026527; AAB82295.1; -
DR GO:0003676; F: nucleic acid binding; IEA.
DR InterPro: IPR004087; KH dom.
DR InterPro: IPR004088; KH type_1.
DR InterPro: IPR00504; RNA_rec_mot.
DR Pfam: PF00013; KH; 4.

```

```

DR Pfam: PF00076; yrm; 2.
DR SMART: SMO0322; KH; 4.
DR SMART: SMO0360; RRM; 2.
DR PROSITE: PS50084; KH type_1; 4.
DR PROSITE: PS50102; RRM; 2.
DR PROSITE: PS00030; RRM RNP 1; FALSE NEG.
SQ SEQUENCE 576 AA; 63271 MW; 01AAE2D1D81C0811 CRC64;

Alignment Scores:
Pred. No.:      8,15e-135      Length:      576
Score:          1951.50      Matches:      388
Percent Similarity: 77.95%      Conservative: 68
Best Local Similarity: 66.32%      Mismatches: 90
Query Match:    33.15%      Indels:      39
DB:              Gaps:      6

US-09-270-437d-8 (1-3283) x 042254 (1-576)

QY      73 ATGAACAAAGCTTTTCAATCGGGAACCTGAGCCCGCCGTCACCGCGACGACCTCGCGAG 132
Db      1 MetAsnIleuTyrIleGlyAsnIleuAsnIleuSerValThrProAlaAspLeuGluIlys 20
QY      133 CTCTTTGGGACAGAGAGAGCTGCCCCGCGGAGAGAGTCTGCTGAAGTCCGGCTTAAGCC 192
Db      21 ValPheAsnAspHisIleuIleSerPheSerGlyGlnPheIleuValIleSerGlyTyrAla 40
QY      193 TTGCTGACATACCCCGACAGAACTGGGCAATCCGCGCATTCGAGACCCCTCTCGGGTAA 252
Db      41 PheValaAspCysProAspGluGlnIleThrAlaMetIleValIleGluThrPheSerGlyIlys 60
QY      253 GTGAAATTCGATGGGAAATATCATGAGTGAAGTATCTACAGTCTCAAAAGCTAAGAGAC 312
Db      61 ValGluIleuHisGlyIleGlnIleuIleGlnIleuIleGlnIleuIleGlnIleuIleGln 80
QY      313 AGGAAATTCGATTCGAAATTCGAAATTCCTCTCACTCACTCACTCACTCACTCACTCACT 372
Db      81 ArgIleIleGlnIleuIleuArgAsnIleProProGlnIleuArgTrpGluValIleuAspGlyIleu 100
QY      373 TTGGCTCAATATGAGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 432
Db      101 LeuAlaGlnIleuTyrThrValGluIleuIleuIleuIleuIleuIleuIleuIleuIleu 120
QY      433 GTTTCACACGTCACTATGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 492
Db      121 ValValaAsnValThrTyrThrAsnArgGlnIleuIleuIleuIleuIleuIleuIleuIleu 140
QY      493 GGGCATCAGTTGAGAACTACTCTTCAAGATTTCTCATCATCCCGATGAAGAGTGAAGC 552
Db      141 GlyHisGlnIleuGlnAsnHisValIleuIleuValSerTyrIleProAspGlnIleu----- 158
QY      553 TCCCTTGGCCCCCTTCAGCGAGCCAGCGGTGGGAGCACTCTTCCCG-----GAG 603
Db      159 SerValGlnIleuProGluIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 178
QY      604 CAAGGCCAC-----GCCCTGGGGGCACTTCAAGCCAGACAGATTAATTTCCCGCTG 657
Db      179 GlnGlySerProValThrAlaGlyAlaProValIleuGlnIleuProValaAspIleProIleu 198
QY      658 CGGATCGGTGTCGCCACCGAGTTGTTGGTGCATCATCGGAAAGAGGCGTTGACCATTA 717
Db      199 ArgIleuIleuValProThrGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 218
QY      718 AAGAACATCATTAAGCAGACCCAGTCCCGGATGATCATCATGAAAGAGAGAGAGAGAGAG 777
Db      219 ArgAsnIleThrIleGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 238
QY      778 GCTGCGAGAGAGAGCTGACCATTCATGACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 837
Db      239 AlaAlaGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 258
QY      838 ATGATCTTGAATCATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 897
Db      259 MetIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 278

```



```

QY 898 CTGAAATCTTGACACACATGGCTTGTTGAGAGCTGATGAAAAAGAGCAGAAAT 957
Db 279 LeuysIleLeuAlaHisAsnAsnPhenValGlyArgLeuIleGlyLysGluIlyArgAsn 298
QY 958 TTGAAGAAATTTGAACATGAAACAGGAGCAGAGATTAACATCTATCTTTGAGGATTTG 1017
Db 299 LeuysIysValGluGlnAspThrGluThrIlySerIleSerSerLeuGlnAspLeu 318
QY 1018 AGCAATATCAACCCGAGAAAGAACCTCACTGCTGAAAGGACAGAGCTGAGCCCTGCCAGT 1077
Db 319 ThrLeuIlyAsnProGluArgThrIleThrValIlySerIleGlnAsnCySerLys 338
QY 1078 GCTGAGATGAGATTAATGAAAGAGCTGAGGCTTGAATGATGATGCTGCTT 1137
Db 339 AlaGluGlnGluIleMetLysLysValArgGluAlaIlyrGluAsnAspValAlaAlaMet 358
QY 1138 AACACCACTCC-----GGATACCTTC----- 1158
Db 359 SerLeuGlnSerHisLeuIleProGlyLeuAsnLeuAlaValGlyLeuPheProAla 378
QY 1159 -----TCCAGCTGTACCCCATCACCAG 1182
Db 379 SerSerAsnAlaValProProProProSerSerValSerGlyAlaAlaProIlyrSerSer 398
QY 1183 TTTGGCCCGTCCGCATCATCTTATCCAGAGCAGAGATTTGAAATCTTCTGATC 1242
Db 399 PheMetPro-----ProGluGlnGluThrValHisValPheIle 411
QY 1243 CCAACCAAGCTGTGGCGCCATCATCGGAGAAAGGGGCGACATCAACAGCTGGCG 1302
Db 412 ProAlaGlnAlaValGlyAlaIleIleGlyLysLysGlnHisIleLysGlnLeuSer 431
QY 1303 AGATTGCGCGAGCTCTATCAAGATGCCCCCTGCGAGAGCCGACGTCGAGGAAAG 1362
Db 432 ArgPheMetAspSerHisIleLysIleAlaProProGluThrProAspSerLysValArg 451
QY 1363 ATGATCATCATCAACCGGCGACCGAGAGCCAGATTCAAGGCCCGAGAGCGATTTGGG 1422
Db 452 MetValValIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyIleTyGly 471
QY 1423 AAAACTGAAAGAGAAAATTCTTTAACCCCAAGAAAGATGAAGCTGGAAGCGCATATC 1482
Db 472 LysLeuLysGlnGluAsnPhePheGlyProLysGlnGluValLysLeuGluThrHisIle 491
QY 1483 AGAGTGCCCTCTCCACAGCTGCGCGGATATTGGCAAGCTGGCAAGCCGTGAAGAA 1542
Db 492 ArgValProAlaSerHisAlaGlyArgValIleGlyLysGlyLysThrValAsnGlu 511
QY 1543 CTGCAAGAACTTAACAGATGAGAGATGATGCTGCTGTCGCAAGCGCCAGATGAAAT 1602
Db 512 LeuGlnAsnLeuThrAlaAlaGluValValValProArgAspGlnThrProAspGluAsn 531
QY 1603 GAGGAAGTATGCTCAAGATTAATCGGCACTTTCTTCTAGCCAGACTGACAGCCGCAAG 1662
Db 532 GluGlnValIleValLysIleIleGlyHisPheTyAlaSerGlnMetAlaGlnIlyAsn 551
QY 1663 ATGAGGGAATTTTACAACAGGTGAAGCCAGCAGAGCAGAAATACCTTCAGGAGTGGCC 1722
Db 552 IleArgAspIleLeuAlaGlnValLysGlnGlnHisGlnLysGlyLysSerGlyLysLeu 571
QY 1723 TCACAGCGCAGCAAG 1737
Db 572 GlnAlaArgArgLys 576

```

```

RESULT 6
Q8CGX0
AC Q8CGX0; PRELIMINARY; PRT; 577 AA.
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 23, Last annotation update)
DE B-actin zipcode binding protein 1.

```

```

OS Rattus norvegicus (Rat)..
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=Sprague-Dawley;
RA Eom T., Singer R.H., Bassell G.J.;
RT "Molecular interactions between rZBP1 and b-actin zipcode required for
RT transport of mRNA and stimulation of spine growth.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF541940; A016210.1;
DR GO; GO:0003676; P:nucleic acid binding; IEA.
DR InterPro; IPR004087; KH_dom.
DR InterPro; IPR004088; KH_type_1.
DR InterPro; IPR00504; RNA_rec_mot.
DR Pfam; PF00013; KH; 4.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00322; KH; 4.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS50084; KH_TYPE_1; 4.
DR PROSITE; PS50102; RRM; 2.
SQ SEQUENCE 577 AA; 63436 MW; 0647676128FPD1BE CRC64;

```

Alignment Scores:

```

Pred. No: 8.03e-134 Length: 577
Score: 1938.00 Matches: 382
Percent Similarity: 79.10% Conservative: 76
Best Local Similarity: 65.98% Mismatches: 95
Query Match: 32.93% Indels: 26
DB: 11 Gaps: 5

```

US-09-270-437D-8 (1-3283) x Q8CGX0 (1-577)

```

QY 73 ATGACAAAGCTTTACATCGGAAACCTGAGCCCGCGTCACCGCCAGCACTCCGCGAG 132
Db 1 MetAsnLysLeuIlyrIleGlyAsnLeuAsnGlnSerValThrProAlaAspLeuGlnLys 20
QY 133 CTCTTTGGGAGCAGAGAGCTGCCCCCTGCGCGGACAGAGTCTGCTGAAAGTCCGCTAGCC 192
Db 21 ValPheAlaGlnIlyHisIleSerIlyrSerGlyGlnPheLeuValLysSerGlyTyAla 40
QY 193 TTGCTGAGCTACCCCGACCAAGAACTGGCGCATTCGCGCATGAGACCTTCGCGGTAA 252
Db 41 PheValAspCySerProAspGlnIlyHisIlyrPalMetLysAlaIleGluThrPheSerGlyLys 60
QY 253 GTGGAATTGATGAGGAAATTCATGGAAGTTGATTACTCACTCTTAAAGCTTAAGAGAC 312
Db 61 ValGlnLeuGlnIlyLysArgLeuGlnIleGlnHisSerValProLysLysGlnArgSer 80
QY 313 AGGAAATTCAGATTGAGAAACATCCCTCTCACTGCGAGTGGAGGTGGATGACTT 372
Db 81 ArgLysIleGlnIleArgAsnIleProProGlnLeuArgThrGluValLeuAspSerLeu 100
QY 373 TTGCTCAATATGAGCAAGTGAAGATGTGGAAACAATGTCACACAGACACAGAACCGCC 432
Db 101 LeuAlaGlnIlyrGlyThrValGlnLeuCysGlnGlnValAsnThrGlnSerGluThrAla 120
QY 433 GTTGTCAAGTCAATATGCAACAGAGAAAGAAAGCAAAATGCCATGAGAAAGCTTAAGC 492
Db 121 ValValAsnValThrTyIserAsnArgGlnGlnThrArgGlnAlaIleMetLysLeuAsn 140
QY 493 GGGCATCAGTTTGAGAACTACTCTTCAAGATTTCTTACATCCCGGATGAAGAGTGAGC 552
Db 141 GlyHisGlnLeuGlnAsnHisAlaLeuLysValSerIlyrIleProAspGlnGlnIleAla 160
QY 553 TCCCTTCGCCCCCTCAGCGAGCCAGCGGTGGAGACACTTTCCCGGAGAGAGCCAG 612
Db 161 -----GlnGlyProGluAsnGlyArgArgGlyGlyPheGlySerArgGlyLysProArg 178
QY 613 -----GCCCTGGGGGCACTTCTCAGCGCAAGCAATGATTTCCGCGTGG 657
Db 179 GlnGlySerProValAlaAlaGlyAlaProAlaLysGlnGlnGlnValAspIleProLeu 198

```

```

QY 658 CGGATCTGCTCCCGACCCAGTTTGTGTCCTCATCGGAAAGAGGGCTTGACCATTA 717
Db 199 ArgLeuLeuValProThrGlnTyrValGlyAlaIleIleGlyLysGlnGlyAlaThrIle 218
718 AAGAATCTCACTAGACAGACCCAGTCCCGGGTAGATATCCATAGAAAAAGAACTCTGGA 777
Db 219 ArgGlnIleThrLysGlnThrGlnSerLysIleAspValHisValGlyGlnAlaGly 238
778 GCTGCAAGAAAGCTGTGCATCCATCCAGCCACCCAGAGGGGACTTCTGAAGATGCCGC 837
QY 239 AlaIleGlnLysAlaIleSerValHisSerThrProGlnGlyCysSerSerAlaCysLys 258
QY 838 ATGATCTTTGAATCATGACAGAAAGGACGATGACCAAACTGACAGAGAGATTCT 897
Db 259 MetIleuGlnIleMetHisLysGlnAlaLysAspThrValAlaAspGlnValPro 278
898 CTGAAATCTTTGGACACACATGCTGCTTGGAGACCTGATTTGAAAAAGAGCAAAAT 957
QY 279 LeuLysIleLeuAlaHisAsnAsnPhenValGlyArgLeuIleGlyLysGlnYargAsn 298
958 TTGAAGAAATTTGAACATGAAACAGGGACCAAGATTAACATCTCTTTGACGATTTG 1017
QY 299 LeuLysLysValGlnGlnAspThrGlnThrLysIleThrIleSerSerLeuGlnAspLeu 318
1018 AGCATATACAAACCCGGAAGAAAGACCTCATCTGGAAGGGACAGTTGAGGCTGTGCCAGT 1077
QY 319 ThrLeuLysAspProGlnGlnArgThrIleThrValLysGlyAlaIleGlnAsnCysGlyArg 338
1078 GCTGAGATGAGATTTATGAGAAGAGCTGCGTGAAGGCTTTGAAATGATATGCTGCTTT 1137
QY 339 AlaGlnGlnGlnIleMetLysLysValArgGlnAlaTyrGlnAsnAspValAlaIleMet 358
1138 AACACCCACTGC-----GGATCTTC----- 1158
QY 359 SerLeuGlnSerHisLeuIleProGlnLysLeuAsnLeuAlaValGlyLeuMetProAla 378
1159 ---TTCAGCTGTACCCCATCACAG-----TTTGGCCCGTTCCCGCAT 1200
QY 379 SerSerSerAlaValProProProSerSerValThrGlyAlaAlaProTyrGlnLys 398
1201 CATCACTCTTATCCAGACAGAGAGATTTGTAATCTTCTCATCCCAACCCAGGCTGTGGC 1260
QY 399 PheMetGlnAlaProGlnGlnGlnMetValGlnValPheIleProIleGlnAlaValGly 418
1261 GCCATCATGCGGAAGAGGGGCAACATCAAAACAGCTGCCAGATTTCCCGGAGCTCT 1320
QY 419 AlaIleIleGlyLysLysGlnHisIleLysGlnLeuSerThrPheAlaSerAlaSer 438
1321 ATCAAGATGCCCCCGGAGAGGCCAGACGTCAGGAAAGATGTCATCATCACCGGG 1380
QY 439 IleLysIleAlaProGlnThrProAspSerLysValArgMetValIleThrGly 458
1381 CCACCGGAAGCCAGTTCAAGGCCCAAGGAGCGATCTTTGGAAACTGAAAGGAAAC 1440
QY 459 ProProGlnAlaGlnIleLysAlaGlnGlyArgIleTyrLysLeuLysGlnGlnAsn 478
1441 TTCTTTAACCCCAAGAAAGAGAGCTGGAAGCCCATTCGAGTGCCTTCTCCACA 1500
QY 479 PhePheGlyProLysGlnGlnValLysLeuGlnThrHisIleArgValProAlaSerAla 498
1501 GCTGCGCGGGTATTTGGCAAAAGTGGCAAGACGTCGACGATGCAAACTTCAACAGT 1560
QY 499 AlaGlyArgValIleGlyLysGlyLysThrValAsnGlnLeuGlnAsnLeuThrAla 518
1561 GCAGAGATCATGCTGCTGTGACCAAAAGCCAGATGAAAAATAGAGAAATGATGTCAGA 1620
QY 519 AlaGlnValValValProArgAspGlnThrProAspGlnAsnAspGlnValIleValLys 538
1621 ATTATGCGGACATTTCTTGTACGACAGCTGCACAGGCAAGATCAGGAAATTTGACAA 1680
QY 539 IleIleGlyHisPheTyrAlaSerGlnMetAlaGlnArgLysIleArgAspIleLeuAla 558

```

```

QY 1681 CAGGTGAAGCAGCAGCAGCAAGAAATACCTTCAGAGAGTGGCTTCACAGCGCAGAG 1737
Db 559 GlnValLysGlnGlnHisGlnLysGlyGlnSerAsnGlnAlaArgArgLys 577

RESULT 7
ID 088477 PRELIMINARY; PRT; 577 AA.
AC 088477;
DT 01-NOV-1998 (TRENBLREL. 08, Created)
DT 01-NOV-1998 (TRENBLREL. 08, Last sequence update)
DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)
DE Coding region determinant binding protein.
GN IGFBP1 OR CRDBP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92217743; PubMed=1559612;
RA Bernstein P.L., Herrick D.J., Prokipcak R.D., Ross J.;
RT "Control of c-myc mRNA half-life in vitro by a protein capable of
RL binding to a coding region stability determinant.";
RN Genes Dev. 6:642-654(1992).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9415886; PubMed=8114742;
RA Herrick D.J., Ross J.;
RT "The half-life of c-myc mRNA in growing and serum-stimulated cells:
RL influence of the coding and 3' untranslated regions and role of
RN ribosome translocation.";
RP Mol. Cell. Biol. 14:2119-2128(1994).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=94179348; PubMed=8132663;
RA Prokipcak R.D., Herrick D.J., Ross J.;
RT "Purification and properties of a protein that binds to the C-terminal
RL coding region of human c-myc mRNA.";
RN J. Biol. Chem. 269:9261-9269(1994).
[4]
RP SEQUENCE FROM N.A.
RX MEDLINE=97322234; PubMed=9178888;
RA Leeds P., Kren B.T., Boylan J.M., Betz N.A., Steer C.J.,
RL Gruppone P.A., Ross J.;
RT "Developmental regulation of CRD-BP, an RNA-binding protein that
RN stabilizes c-myc mRNA in vitro.";
RL Oncogene 14:1279-1286(1997).
[5]
RP SEQUENCE FROM N.A.
RA Ross J., Prokipcak R.D., Leeds P., Doyle G.A.R., Betz N.A.,
RL Fleisig A.J.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
[6]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Embryo;
RL MEDLINE=2108560; PubMed=11217851;
RA Kawai U., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Araiawa T., Hara A., Fukumishi Y., Komio H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Giesi C., King B., Kochia H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staahl F., Suzuki R., Tomita M., Wagner U., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni U., Maehima U., Nazarelli I., Sakamoto N.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakuma K.-F.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohseuki S.,

```

RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
CC -I- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR EMBL; AF061569; AAC72743.1; -.
DR EMBL; AK013940; BAB29071.1; -.
DR HSSP; P11940; 1CVD.
DR MGD; MGI:1890357; Igf2bp1.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR004087; KH dom.
DR InterPro; IPR004088; KH type_1.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00013; KH; 4.
DR Pfam; PF00076; xtm; 2.
DR SMART; SM00322; KH; 4.
DR SMART; SM00360; RM; 2.
DR PROSITE; PS00084; KH_TYPE_1; 4.
DR PROSITE; PS0102; RNM; 2.
SQ SEQUENCE 577 AA; 63450 MW; EFBBIAF2PF9F0344 CRC64;

Alignment Scores:
Pred. No.: 1,13e-133 Length: 577
Score: 1936.00 Matches: 381
Percent Similarity: 79.10% Conservative: 77
Best Local Similarity: 65.80% Mismatches: 95
Query Match: 32.89% Indels: 26
DB: Gaps: 5

US-09-270-437d-8 (1-3283) x 088477 (1-577)

QY 73 ATGAACAAGCTTTATCATCGGGAGAACCCCGCCCTGACCGCCGACGACCTCCGCGAG 132
Db 1 MetAsnIysLeuTYrIleGlyAsnLeuAsnIuSerValThrProAlaAspLeuGluLys 20

QY 133 CTCTTTGGGACAGAGAACTGCCCCCTGGCGGACAGAGTCTGCTGAAGTCGGGTACGCC 192
Db 21 ValPheAlaGlnHisLysIleSerTySerGlyInPheLeuValLysSerGlyTyrAla 40

QY 193 TTGGTGAGACTACCCCGACCGAAGTGGGACATCCGGCCATCCGAGACCTTCGGGTAAA 252
Db 41 PheValAspCysProAspGlnHisStrpAlaMetLysAlaIleGluThrPheSerGlyLys 60

QY 253 GTGGAATTGCATGGGAAATCATGGAAGTGTATTACTGCTTAAAGAACTGAAGAGC 312
Db 61 ValGluLeuGlnGlyLysArgLeuGluIleGlnHisSerValProLysGlnArgSer 80

QY 313 AGGAAATTTGAGATTGCAAAACATCCTCTCTCACTGACAGTGGAGGTGGATGACTT 372
Db 81 ArgLysIleGlnIleArgAsnIleProProGlnLeuArgTyrGluValLeuAspSerLeu 100

QY 373 TTGGCTCAATATGAGGACAGTGGAGAAATGGAACAATGCAACAACAGACAGAAACCGCC 432
Db 101 LeuAlaGlnIlyrGlyThrValGlnAsnCysGlnGlnValAsnThrGlnSerGluThrAla 120

QY 433 GTTGTCAAGCTCATATGCAACAAGAGAAGCAAAATAGCCATGAGAGCTAAGC 492
Db 121 ValValAsnValThrTyrSerAsnArgGlnGlnThrArgGlnAlaIleMetLysLeuAsn 140

QY 493 GGGCATCAGTTTGAAGAACTACTCTTCAAGATTCTTCAATCCCGATGAAGAGGTAGC 552
Db 141 GlyHisGlnLeuGlnHisAlaLeuLysValSerTyrIleProAspGlnIleThr 160

QY 553 TCCCTTCGCGCCCTCAGAGGCCAGCGTGGGGACACACTCTTCCCGGAGCAAGCCAC 612
Db 161 -----GlnGlyProGlnGlnGlyArgGlyGlyPheGlySerArgGlyGlnProArg 178

QY 613 -----GCCCTGGGGCACTTCTCAGGCGACAGATGATTTCCCGCTG 657
Db 179 GlnGlySerProValAlaAlaGlyAlaProAlaLysGlnGlnProValaAspIleProLeu 198

QY 658 CGGATCCTGGTCCCCACCCGTTGTTGGTCCATCATCGAAAGAGGGCTTGACATA 717
Db 199 ArgLeuLeuValProThrGlnTyrValGlyAlaIleIleGlyLysGlnGlyAlaThrIle 218

QY 718 AAGAACATCATTAACCAAGACCCAGTCCCGGTAATATCCATGAAAAAGAACTGTGA 777
Db 219 ArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgLysGlnHisAlaGly 238

QY 778 GCTGAGAGAGCCCTGACCAATCCATGCCACCCCGAGGGGACTTCTGAAGATGCCG 837
Db 239 AlaAlaGlnLysValaIleSerValHisSerThrProGlnGlyCysSerSerLysLys 258

QY 838 ATGATTTCTTGAATCATGCAGAAAGAGGACAGATGAGACCAACTGAGCGAAGATGCT 897
Db 259 MetIleLeuGluIleMetHisLysGlnAlaLysAspThrLysThrAlaAspGluValPro 278

QY 898 CTGAAATCTTGGCACACAAATGCGCTTGGTGAAGACTGATTTGAAAAAGAGCAGAAAT 957
Db 279 LeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGlnGlyArgAsn 298

QY 958 TTGAAGAAATTTGAACAGAAACAGAGGACCAAGATTAACAATCTACTTTGACGAGATTG 1017
Db 299 LeuLysLysValaGlnGlnAspThrGluThrLysIleThrIleSerSerLeuGlnAspLeu 318

QY 1018 AGCATATTAACAACCCGAAAGAACCATCATCTGTGAAGGCGACAGTTGAGCGCTGCCAGT 1077
Db 319 ThrLeuTyrAsnProGlnLysArgThrIleThrValLysGlyAlaIleGlnAsnCysArg 338

QY 1078 GCTGAGATTAAGATTATGAGAGAGCTGCGTGAAGCTTTGAAAATGATATGCTGCTGTT 1137
Db 339 AlaGlnGlnGlnIleMetLysLysValaArgGlnLysLysValaGlnAsnAspValAlaAlaMet 358

QY 1138 AACACCCCACTCC-----GGATACTTC----- 1158
Db 359 SerLeuGlnSerHisLysIleProGlyLeuAsnLeuAlaAlaValGlyLeuPheProAla 378

QY 1159 ---TCCAGCTGTATCCCCCATCACAG-----TTGGCCCGTTCGCCGAT 1200
Db 379 SerSerSerAlaValProProProProSerSerValThrGlyAlaAlaProTyrSerSer 398

QY 1201 CATCACTCTTAATCCAGAGCAGAGAGATTGTGAATCTTTCATCCCAACCCAGCGCTGGGC 1260
Db 399 PheMetGlnAlaProGlnGlnGlnMetValGlnValPheIleProAlaGlnAlaValGly 418

QY 1261 GCCATCATCGGGAAGAGAGGGGGCACACATCAACAGCTGGGAGATTGGCGGACTCT 1320
Db 419 AlaIleIleGlyLysLysGlyGlnHisLysLysGlnLeuSerArgPheAlaSerAlaSer 438

QY 1321 ATCAAGATTGCCCTCGGAGAGCCCGACAGCTGACGCAAGAAAGATGATCATCATCACCGGG 1380
Db 439 IleLysIleAlaProProGluThrProAspSerLysValaArgMetValaIleThrGly 458

QY 1381 CCACCGGAGCCCAAGTTCAAGGCCCGCAGGAGCGGATCTTTGGGAAACTGAAGAGAAAC 1440
Db 459 ProProGlnLysAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysLeuLysGlnGlnAsn 478

QY 1441 TTCTTTAACCCTCAAGAAAGAGTGAAGCTGGAAGCGCATATCAGATGCGCTCTTCACA 1500
Db 479 PhePheGlyProLysGlnGlnValaLysLeuGlnThrHisIleArgValProAlaSerAla 498

QY 1501 GCTGCGCGGCTGATTGGCAAGGTTGGCAAGACCGTGAACAACGACGCAACTTAACAGT 1560
Db 499 AlaGlyArgValaIleGlyLysGlyLysThrValaAsnGlnLeuGlnHisLeuThrAla 518

QY 1561 GCAGAAATCATGCTGCTGCTGTGAACCAACCGCATGAATGAATGAAGAAATGATCGTACA 1620
Db 519 AlaGlnValaValaValaProArgAspGlnThrProAspGlnHisAspGlnValaIleValLys 538

QY 1621 ATTATCGGGGACTCTTGTGACCGCAGATGTCGACACCGCAAGATCAGGAAATTTGTACAA 1680
Db 539 IleIleGlyHisPheTyrLysSerGlnMetAlaGlnArgLysIleArgAspIleLeuAla 558

QY 1681 CAGGTGAACAGCAGAGAGAAATACCTCAGGAGAGTGGCTTCACAGCGCAGCAAG 1737
Db 559 GlnValLysGlnGlnHisGlnLysGlnSerLysSerAsnLeuAlaGlnAlaArgArgLys 577

RESULT 8
Q9NZ18 PRELIMINARY; PRT; 577 AA.
AC Q9NZ18;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE mRNA-binding protein CRBP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ioannidis P., Trangas T., Dimitriadis E., Samiotaki M.,
RA Panoutsakopoulos G., Kyriazoglou I., Voutzoulis S., Tsiapalis C.M.,
RA Kitis C., Agnantis N., Pandis N.;
RT "Ectopic expression of a KH-domain containing protein, highly
homologous to both human IMP-1 and mouse CRD-BP, in benign and
malignant mesenchymal tumors."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR EMBL; AF198254; AAF37203.1; -.
DR HSSP; P1340; 1CVJ.
DR GO; GO:0003676; P:nucleic acid binding; IEA.
DR InterPro; IPR004087; KH_dom.
DR InterPro; IPR004088; KH_type_1.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00013; KH; 4.
DR Pfam; PF00076; rim; 2.
DR SMART; SM00322; KH; 4.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS50084; KH_TYPE_1; 4.
DR PROSITE; PS50102; RRM; 2.
DR SEQUENCE 577 AA; 63456 MW; 0749A060F252D81D CRC64;
SQ
Alignment Scores:
Pred. No.: 133e-133 Length: 577
Score: 1935.00 Matches: 382
Percent Similarity: 78.93% Conservative: 75
Best Local Similarity: 65.98% Mismatches: 96
Query Match: 32.87% Indels: 26
Gaps: 5
US-09-270-437d-8 (1-3283) x Q9NZ18 (1-577)
QY 73 ATGAAACAGCTTTACATCGGAACTGAGCCCGCGTCACGCGCAGACACTCCGGCAG 132
Db 1 MetAsnLysLeuThrIleGlyAsnLeuAsnGluSerValThrProAlaAspLeuGluLys 20
QY 133 CTCTTTGGGAGACGGAAGCTGCCCCGCGGGGACAGTCTCTGTAAGTCCGGCTACGCGC 192
Db 21 ValPheAlaGluHisLysIleSerLysSerGlyGlnPheLeuValLysSerGlyTyrAla 40
QY 193 TTCTGGAGACACACCGCCGACGAGCTGGCCATCGAGCCCTCTCGGGTAA 252
Db 41 PheValAspCysProAspGlnHisLysPheAlaMetLysAlaIleGlnThrPheSerGlyLys 60
QY 253 GTGGAATTGCATGGGAAATCATGGAATTGATTACTCACTCTCTTAAAGTCTAAGAGC 312
Db 61 ValGlnLeuGlnLysArgLeuGlnIleGlnHisSerValProLysLysGlnArgSer 80
QY 313 AGGAAATTCAGATTGGAACATCCCTCCACCTGAGAGGGAGGTGTGATGACT 372
Db 81 ArgLysIleGlnIleArgAsnIleProProGlnLeuArgTyrGlnValLeuAspSerLeu 100
QY 373 TTGGCTCAATATGAGGACAGTGAATGTGAAACAAGTCAACAGACAGAGAAACCGCC 432
Db 101 LeuAlaGlnThrValGlnThrValGlnAsnGlyGlnValAsnThrGlnLysGlnThrAla 120
QY 433 GTTGTCACCTCATATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 492
Db 121 ValValAsnValThrTyrSerAsnArgGlnGlnThrArgGlnAlaIleMetLysLeuAsn 140

QY 493 GGCATCAGTTTGAAGAACTACTCTTCAAGATTCTTACATCCGGATGAAAGAGTGAC 552
Db 141 GlnHisGlnLeuGlnAsnHisAlaLeuLysValSerTyrIleProAspGlnGlnIleAla 160
QY 553 TCCCTTCGCCCCCTCAGCCGAGCCAGGTGGGACACACTTCCCGGACAGAGCCAC 612
Db 161 -----GlnGlnAsnGlyArgArgGlyGlyPheGlySerArgGlyGlnProArg 178
QY 613 -----GCCCTGGGGCACTTCTCAGAGCCAGACAGATTGATTTCCCGGTG 657
Db 179 GlnGlySerProValAlaAlaGlyAlaProAlaLysGlnGlnValAspIleProLeu 198
QY 658 CGGATTCCTGATCCCGACCCAGCTTTGTTGGTCCATCATCGAAGAGAGGCTTGACCTA 717
Db 199 ArgLeuLeuValProThrGlnThrValGlnAlaIleIleGlyLysGlnGlyAlaThrIle 218
QY 718 AAGAAATCATCACTAAGCAGACCCAGTCCGGGTATGATTCATAGAAAGAGAACTCTGA 777
Db 219 ArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgLysGlnAsnAlaGly 238
QY 778 GCTGAGAGAGAGCTGTGACCATCATGCTCAGCCAGAGGGAGCTTCAAGCATGCGCG 837
Db 239 AlaAlaGlnLysAlaIleSerValHisSerThrProGlnGlyCysSerSerAlaCysLys 258
QY 838 ATGATTTCTTGAATCATGACAGAAAGGCAATGACACCAATCTTGCCGAGAGATTCT 897
Db 259 MetIleLeuGlnIleMetHisLysGlnLysAlaLysAspThrLysThrAlaAspIleValPro 278
QY 898 CTGAAATCTTGGCACACAAATGCTTGGTGAACATGATGAAAGAGAGGCGAGAAAT 957
Db 279 LeuLysThrLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGlnGlyArgAsn 298
QY 958 TTGAAGAAATTTGAACATGAAACAGGACCAAGATTAACATCTCTTGGAGAGATTGG 1017
Db 299 LeuLysLysValGlnGlnAspThrGlnThrLysIleThrIleSerSerLeuGlnAspLeu 318
QY 1018 AGCATATACACCCGAGAAAGAACATCACTGTGAAGGGCAGCAAGTGGAGCTGTGCCAAT 1077
Db 319 ThrLeuTyrAsnProGlnLysThrIleThrValLysGlyAlaIleGlnLysnGlyCysArg 338
QY 1078 GCTGAGATGAGATTTATGAAAGAGCTGGGTGAGGCTTGAATGATGCTGGTGT 1137
Db 339 AlaGlnGlnGlnIleMetLysLysValArgGlnAlaLysGlnLysnAspValAlaAlaMet 358
QY 1138 AACACCCACTCC-----GGATTACTTC----- 1158
Db 359 SerLeuGlnSerHisLeuThrProGlyLeuAsnLeuAlaAlaValGlyLeuPheProAla 378
QY 1159 ---TCCAGCTGTACCCCATCAG-----TTGGCCCGCTCCCGCAT 1200
Db 379 SerSerSerAlaValProProProProSerSerValThrGlyAlaAlaProLysSerSer 398
QY 1201 CATCACTTTATCCAGAGAGAGAGATGGAATCTCTTCATCCCAACCGAGGTGGGC 1260
Db 399 PheMetGlnAlaProGlnGlnGlnMetValGlnValPheIleProAlaGlnAlaValGly 418
QY 1261 GCCATCATCGGAGAGAGAGGAGGACATCAACAGCTGGCAGATTCGCGGAGCTCT 1320
Db 419 AlaIleIleGlyLysLysGlnHisLysIleGlnLeuSerArgPheAlaSerAlaSer 438
QY 1321 ATCAAGATTGCCCTGGGAGAGCCCAAGACGTACAGGAAAGATGTGATCATCACCGGG 1380
Db 439 IleLysIleAlaProProGlnThrProAspSerLysValArgMetValIleIleThrGly 458
QY 1381 CGACCGGAGAGCCAGTTCAAGGCGCCAGGAGCGATCTTTGGAGAACTGAAAGAGAGAAAC 1440
Db 459 ProProGlnLysAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysLeuLysGlnLysn 478
QY 1441 TTCTTTAACCCCAAG 1500
Db 479 PhePheGlyProLysGlnGlnValLysLeuGlnThrHisLysArgValProAlaSerAla 498

```
QY 1501 GCTGCGCGGCTGATTGGCAAGGTGGCAAGACCGTGAACGAACCTGCAAGAACTTAACCACT 1560
Db 499 AAGYArYgValIleelYysgIyLysThrValAsnGluLeuEndInleuNthra 518
QY 1561 GCAGAAGTCATGCTGCTGCTGACCAAGCCAGATGAAATGAGAAATGAGAAATGATCGTACA 1620
Db 519 AAGlValValValProArgAspGlnThrProAspGluAsnAspGlnValIleVallys 538
QY 1621 ATTATCGGCACTCTTCTTCTGCTGACCACTGCAACGCGCAGAGATCAAGGAAATTTGACAA 1680
Db 539 IleIleGlyHisPheTyRAlaSerGlnMetAlaGlnAspGlyIleArgAspIleuAla 558
QY 1681 CAGGTGAAGCAGCAGAGCAGCAAAATACCTCAGGAGTGCCTCAGACAGCAGCAAG 1737
Db 559 GlnVallysgIInGlnHisGlnLysGlyGlnSerAsnGlnAlaGlnAlaArgArglys 577
RESULT 9
QBBRH1
ID QBBRH1 PRELIMINARY; PRT; 577 AA.
AC QBBRH1;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Insulin-like growth factor 2.
GN IGF2BP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK044850; BAC32119.1; -.
DR MGB; MGI:1890357; IGF2BP1.
DR GO; GO:0003676; P.nucleic acid binding; IEA.
DR InterPro; IPR004087; KH dom.
DR InterPro; IPR004088; KH type_1.
DR InterPro; IPR005054; RNA_rec_mot.
DR Pfam; PF00013; KH; 4.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00322; KH; 4.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS50084; KH_TYPE_1; 4.
DR PROSITE; PS50102; RRM; 2.
SQ SEQUENCE 577 AA; 63378 MW; D75897450841532E CRC64;
Alignment Scores:
Pred. No.: 3.69e-133 Length: 577
Score: 1929.00 Matches: 380
Percent Similarity: 78.93% Conservative: 77
Best Local Similarity: 65.63% Mismatches: 96
Query Match: 32.77% Indels: 26
DB: 11 Gaps: 5
US-09-270-437D-8 (1-3283) x QBBRH1 (1-577)
QY 73 ATGAACAAGCTTATTCGGAACCTGAGCCCGCGCTCACCCCGACAGCACTTCGGCAG 132
Db 1 MetAsnIysIeuYrIleGlyAsnIleuAsnGlnSerValThrProAlaAspIleuGlnlys 20
QY 133 CTCCTTGGGAGCAGGAAGCTGCCCTGCGCGGACAGAGTCTCTGTAAGTCGGGCTAGCC 192
Db 21 ValPheAlaGlnHisIleSerTyRserGlyGlnPheIeuVallysserGlyTyRAla 40
QY 193 TTGCTGACATCAACCCAGCAAGAACTGGGCGCATCGCGCATGAGAACCTCTCGGGTAA 252
Db 41 PheValAspCysProAspGlnHisThrPAlaMetIysAlaIleGlnThrPheSerGlylys 60
```

```
QY 253 GTGAATTCATGGGAAATATCATGGAATTGATTACTAGTCTTAAAAAGCTTAAGAGC 312
Db 61 ValGlnLeuGlnGlyLeuArgIleuGlnIleGlnHisSerValProLysGlnArgSer 80
QY 313 AGGAAATTCAGATTCGAAACATCCCTCTCACCTCGAGCTGGGAGAGTGTGATGACTT 372
Db 81 ArgIysIleGlnIleArgAsnIleProGlnLeuAspGlyValIleuAspSerIeu 100
QY 373 TTGGCTCAATATGGACAGTGGAGATGTGAAACAAGTCAACACAGACACAGAAACCGCC 432
Db 101 LeuAlaGlnTyRgIyThrValGlnAsnGlyGlnGlnIleValAsnThrGlnSerIuThra 120
QY 433 GTTGCAACGTTCACATATGCAACAAGAAAGCAAAATATGCCATGAGAGAGCTTAAC 492
Db 121 ValValAsnValThrTyRserAsnArgGlnGlnThrArgGlnAlaIleMetLysIeuAsn 140
QY 493 GGGCATCAGTTTGAAGAACTACTCCTTCAGAAATTTCTATCATCCCGATGAGAGGTAGC 552
Db 141 GlyHisGlnLeuGlnAsnHisAlaLeuLysValSerTyRileProAspGlnGlnIleThr 160
QY 553 TCCCTTCGCCCCCTCAGACGAGCCAGCGTGGGACCACTCTCCCGGAGCAGAGCCAC 612
Db 161 -----GlnGlyProGlnAsnGlyArgArgGlyGlyPheGlySerArgGlyGlnProArg 178
QY 613 -----GCCCTGGGGGACCTTCTCAGGCGCAGACAGATTTCCCGCTG 657
Db 179 GlnGlySerProValAlaAlaGlyAlaProAlaLysGlnGlnProValAspIleProIeu 198
QY 658 CGGATCCGTCGCGCCACCCAGTTTGTGTCCTCATTCGGAAGAGAGGCTTACCACTA 717
Db 199 ArgIeuIeuValProIleGlnTyRValGlyAlaIleIleGlyLysGlnGlyAlaThrIle 218
QY 718 AAGAATCATCAAGACAGACCCAGTCCCGGATAGATATCCATGAAAGAAAGAACTCTGGA 777
Db 219 ArgAsnIleThrIysGlnThrGlnSerLysIleAspValHisArgLysGlnAsnAlaGly 238
QY 778 GCTCAGAGAAAGCTGTACCATCATGCCATCCCAAGAGGAGGACTTGAAGCATGCCGC 837
Db 239 AlaAlaGlnLysAlaIleSerValHisSerThrProGlnGlyCysSerSerAlaCysLys 258
QY 838 ATGATTTCTGAATATCATGACAGAAAGAGCAGATGACACCAACTAGCCGAAGAAATTCCT 897
Db 259 MetIleuGlnIleMetHisLysGlnAlaLysAspThrIysThrAlaAspGlnValPro 278
QY 898 CTGAATATCTTGGACACAATAGGCTGTGTAAGATGTAATGGAAGAAAGAGCAAAAT 957
Db 279 LeuLysIleLeuAlaHisAsnAsnPheValGlyArgIleuIleGlyLysGlnGlyArgAsn 298
QY 958 TTGAAGAAAATTGAACATGAAACAGGAGCAAGATTAACAATCTCATCTTTGACAGATTGG 1017
Db 299 LeuLysLysValGlnGlnAspThrGlnThrLysIleThrIleSerSerLeuGlnAspIeu 318
QY 1018 AGCATATACAAACCCGGAAGAAACATCATCTGTGAAGGAGCAGTTGAGGCTTGCCAGT 1077
Db 319 ThrLeuIysAsnProGlnArgGlnThrIleThrValLysIleValAlaIleGluAsnGlyCysArg 338
QY 1078 GCTGAGATAGAGATTATGAAGAGCTGCGTGAAGGCTTGAATAATGATATGCTGCTGT 1137
Db 339 AlaGlnGlnGlnIleLeuMetLysValArgGlnAlaTyRLeuAsnAspValAlaAlaMet 358
QY 1138 AACACCACTCC-----GGATACTTC----- 1158
Db 359 SerLeuGlnSerHisIleuIleProGlyLeuAsnIleuAlaValGlyLeuPheProAla 378
QY 1159 ---TCCAGCCTGAACCCCATCACAG-----TTGGCCCGCTTCCGCAT 1200
Db 379 SerSerSerAlaValProProProProSerSerValThrGlyAlaAlaProTyRSerSer 398
QY 1201 CATCACTTATTCAGAGCAGAGATGTGAATCTTCAATCCCAACCCAGCTGAGGCG 1260
Db 399 PheMetGlnAlaProIleGlnGlnMetValGlnValPheIleProAlaGlnAlaValGly 418
```



```

Db 239 AlaAlaGluValAlaIleSerValHisSerThrProGluGlyCysSerSerAlaCysLys 258
QY 838 ATGATTTCTTGAATCATGATCAGAGGACAGATGACCAATCTGACCAAGAGAGATTCCT 897
Db 259 MetIleLeuGluIleMetHisLysGluAlaLysAspThrLysThrAlaAspGlyValPro 278
QY 898 CTGAAATCTTGGACACACATGGCTTGGTGGAGACCTGATTTGGAAGAAAGAGCGAGAAT 957
Db 279 LeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsn 298
QY 958 TTGAAGAAATTTGAATGAAATGAAACGAGGACCAAGATACATCTCATCTTTGACAGATTG 1017
Db 299 LeuLysLysValGluGlnAspThrGlnThrLysIleThrIleSerSerLeuGlnAspLeu 318
QY 1018 AGCATATTCACACCGGAAAGAACCATCATCTGTAAGGGGACAGTGGGCTGTGCCAGT 1077
Db 319 ThrLeuLysAsnProGluArgThrIleThrValLysGlyAlaIleGluAsnCysCysArg 338
QY 1078 GCTGAGATAGAGATTATGAAAGACCTGCGTAGGCTTTGAAATGATATGCTGCTGTT 1137
Db 339 AlaGluGlnGluIleMetLysLysValArgGluAlaLysArgLysAsnValAlaAlaMet 358
QY 1138 AACGCCACATCC-----GGAATCTTC----- 1158
Db 359 SerLeuGlnSerHisLeuIleProGlyLeuAsnLeuAlaAlaValGlyLeuPheProAla 378
QY 1159 -----TCCAGCTGTACCCCATCACCAG-----TTGGCCCGCTTCCCGCAT 1200
Db 379 SerSerSerLeuAlaValProProProProSerSerValThrGlyAlaAlaProLysSer 398
QY 1201 CATCACTCTTATCCAGAGACAGAGATTTGATCTTCTTCAATCCCAACCCAGCGCTGGGC 1260
Db 399 PheMetGlnAlaProGluGlnGluMetValGlnValPheIleProAlaGlnAlaValGly 418
QY 1261 GCATCATCTGGGAAGAAAGGGGGACACATCAACAGCTGGCGAGATTGCGCGAGCTCT 1320
Db 419 AlaIleIleGlyLysLysGlyGlnHisIleLysGlnLeuSerArgPheAlaSerAlaSer 438
QY 1321 ATCAAGATTGCGCTGGCGAAGGCCAGAGCTGACGCAAGCAAGTGTCATCATCACCGGG 1380
Db 439 IleLysIleAlaProProGluThrProAspSerLysValArgMetValIleThrGly 458
QY 1381 CCACCGGAGCCAGTTCACAGGCCACAGGACGATCTTTGGGAACTGAGAAAGAGAAAC 1440
Db 459 ProProGluAlaGlnPheLysAlaGlnGlyArgIleLysLysLysLeuLysGluGln 478
QY 1441 TTCTTTAAACCCAAAGAGAGAGTGAAGCTGGAACCGCATATCAGATGCCCTTTCACA 1500
Db 479 PhePheGlyProLysGluGluValLysLeuGlnThrHisIleArgValProAlaSerAla 498
QY 1501 GCTGGCCGGGTGATTTGGCAAAAGGTGGCAAGACCGTGAACAACTGCAAGAACTTAACCA 1560
Db 499 AlaGlyArgValIleGlyLysGlyGlyLysThrValAsnGlnLeuGlnAsnLeuThrAla 518
QY 1561 GCAGAAAGTCATCGTGCCTGTCGACCAAGCCAGATGAAATGAGAGAGAGATGCTGACAG 1620
Db 519 AlaGluValValValProArgAspGlnThrProAspGluAsnAspGlnValIleValLys 538
QY 1621 ATTATGGGCACTTCTTTGCTAGCCAGATGCAACAGCGCAAGATCAGGAAATTTGACA 1680
Db 539 IleIleGlyHisPheLysAlaSerGlnMetAlaGlnArgLysIleArgAspIleLeuAla 558
QY 1681 CAGGTGACAGACAGAGAGAAATACCTCAGGAGGTGGCCCTCAACAGGCGACAG 1737
Db 559 GlnValLysGlnGlnHisGlnLysGlyGlnSerAsnLeuAlaGlnAlaArgLys 577

```

```

DE Vg1 RNA binding protein (Decapentaplegic and Vg-related 1, RNA binding
DE protein).
GN DVALRBP.
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang Q., Yaniv K., Oberman F., Wolke U., Gilt A., Fromer M.,
RA Taylor W., Meyer D., Standart N., Raz E., Yisraeli J.K.,
RT "Vg1 RBP intracellular distribution and evolutionarily conserved
RT expression suggest multiple roles during development.";
RL Mech. Dev. 0:0-0(1999).
[2]
RP SEQUENCE FROM N.A.
RA STRAIN=AB; TISSUE=Body;
RC MEDLINE=2388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshimiki S., Carninci P., Pirange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson J.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huljk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalka U., Smalusz D.E., Schnerch A., Schein J.E.,
RT Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=AB; TISSUE=Body;
RC Strausberg R.;
RA Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR EMBL: AF161270; AAD45610.1; -
DR EMBL: BC045873; AAH45873.1; -
DR ZFIN: ZDB-GENE-000308-1; dvrlbp.
GO GO:0003676; R:nucleic acid binding; IEA.
DR InterPro: IPR004087; KH_dom.
DR InterPro: IPR004088; KH_type_1.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00013; KH; 4.
DR DR: pfam: PF00076; rrm; 2.
DR SMART: SM00322; KH; 4.
DR SMART: SM00360; RRM; 2.
DR SMART: PS50084; KH_type_1; 4.
DR PROSITE: PS0102; RRM; 2.
DR PROSITE: PS0102; RRM; 2.
SQ SEQUENCE 582 AA; 63351 MW; 9DAE63200681B306 CRC64;

Alignment Scores:
Pred. No.: 2,52e-130 Length: 582
Score: 1890.50 Matches: 377
Percent Similarity: 78.63% Conservative: 72
Best Local Similarity: 66.02% Mismatches: 97
Query Match: 32.12% Indels: 25
DB: 13 Gaps: 5

US-09-270-437D-8 (1-3283) x Q9PM80 (1-582)
QY 73 ATGAACAAGCTTATACATCGGAAGCTGAGCCCGCGCTGACGCGCGAGAGCTCGGCGAG 132
DB 1 MetAsnLysLeuTyrlleGlyAsnValSerGluGlnAlaSerAlaLeuAspLeuGluSer 20

```


QY 133 CTCCTTTGGGAGACAGAAAGCTGCCCTGGCGGAGACAGTCTCTGTAAGTCCGGCTAGACC 192
 Db 21 ILePheGluGlnThrPylsIleProPheSerAlaProPheLeuValLysSerGlyTyrAla 40
 QY 193 TTGCGTGAACATCCCGGACCAACAACTCCGGCCATCCGGCATCGAGACCCCTCGGGTAA 252
 Db 41 PheValAspCysProAspGluLysValAlaMetArgAlaIleAspThrLeuSerGlyLys 60
 QY 253 GTGGAATTCAGTGGGAAATATCATGGAAGTTGATTACTGCTCTTAAAGCTAAGAGC 312
 Db 61 ValGluLeuHISGlyLysValLeuGluValAlaGluHisSerValProLysArgGluHisSer 80
 QY 313 AGGAAATTCAGATTCCAAACATCCCTCTCAACCTCGAGCTGGGAGGTGGTGGAGACTT 372
 Db 81 CysLysLeuGlnIleArgAsnIleProProHISmetGlnTrpGluValLeuAspGlyLeu 100
 QY 373 TTGGCTCAATATGGGACATGAGAAATGTGGAACAAGTCAACACAGACAGAAACCCGC 432
 Db 101 LeuAlaGlnTyrGlyThrValGluSerCysGluGlnValAsnThrAspThrGluThrAla 120
 QY 433 GTTGTCACAGTCATATATGCAACAGAGAGAGAAATAATGCCATGAGAGAGCTTAAGC 492
 Db 121 ValValAsnValArgTyrGlyValAlaLysAspGlnAlaArgGluAlaMetAspLysLeuAsn 140
 QY 493 GGGCATCAGTTTGAGAACTACTCTTCAAGATTCTCTCATCCCGGATGGAAGAGGTGAGC 552
 Db 141 GlyPheLeuMetGluAsnTyrAlaLeuLysValSerTyrIleProAspGlnThrAlaAla 160
 QY 553 TCCTCTTGGCCCTCAGCAGAGCCAGCGTGAG-----GACCACTCTTCCGGAGCA 606
 Db 161 AlaAspAlaProAlaValAlaGlyLysArgArgGlyPheAsnProArgGlyProArgArgGln 180
 QY 607 GGCCACGCCCTGGGGGCACTTTCAGGCCACAGAGATTGATTCCTCCGCTGGGATCCTG 666
 Db 181 GlySerProSerLeuGlyValaArgProLysLeuGlnSerAspValProLeuArgLeuLeu 200
 QY 667 GTCCCAACCCAGTTGTTGTGTGCCATCATCCGGAAGGAGGGCTTACCACTAAGAACATC 726
 Db 201 ValProThrGlnPheValAlaIleGlyLysGluGlyValaThrIleArgAsnIle 220
 QY 727 ACTAGCAGACCCAGTCCCGGTAGATATCATTAAGAAAGAACTGTGAGCTGACAG 786
 Db 221 ThrLysGlnThrHisSerLysIleAspIleHisArgLysGluAsnAlaGlyAlaGlu 240
 QY 787 AAGCTGTCAACCATCATGACACCCAGGAGGACTTTCGAAGCATGCCGATGATTTCT 846
 Db 241 LysProIleThrValHisSerThrProGluGlyCysSerSerAlaCysArgAsnIleMet 260
 QY 847 GAAATATGACAGAAAGAGCAGATGAGACCAACTAGCCGAGAGATTCCTCGAATAATC 906
 Db 261 GlnIleMetGlnLysIleAlaIleAspThrLysIleThrGluGluIleProLeuLysIle 280
 QY 907 TTGGCAGACAAATGGCTGTGTTGGAAGACTGATTAAGAAAGAGCAATTTGAAGAA 966
 Db 281 LeuAlaHisAsnAsnPheValAlaGlyArgLeuIleGlyLysGluGlyArgAsnLeuLys 300
 QY 967 ATTGAACATGAAACAGGAGCAAGATTAACATTCATCTTTGGAGATTGGACATATAC 1026
 Db 301 IleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnAspLeuThrLeuTyr 320
 QY 1027 AACCCGGAAGAAACCATCTGCTGAGAGGAGACAGTGGCCGCTGCGATGCTGAGATA 1086
 Db 321 AsnProGlnArgThrIleThrValLysGlyThrLeuAspAlaCysAlaLysAlaGluGlu 340
 QY 1087 GAGATTATGAAAGAGCTGCGTGGAGCGCTTGAATAATGATATGCTGCTGTTAAACCCAC 1146
 Db 341 GlnIleMetLysLysValaArgLysTyrGluAsnAspValaAlaIleMetHisLeuGln 360
 QY 1147 TCC-----GGATCTTC----- 1158
 Db 361 SerAsnLeuIleProGlyLeuAsnLeuAlaLeuGlyLeuPheProGlyAlaAlaSer 380

QY 1159 TCAGCCCTGTAAACCCCATCAACCAAGTTTGCCCGTTCG-----CATCATCAC 1206
 Db 381 GlyLysIleSerProSerValAlaSerGlyProProProGlyAlaGlnAlaGlyTyrGln 400
 QY 1207 TCTTAT-----CGAGCAGAGAGATTGTGAATCTCTTCAATCCCAACCAAGCTGTG 1257
 Db 401 SerPheGlyAlaGlnMetGluSerGlnThrValHisLeuPheIleProAlaLeuAlaVal 420
 QY 1258 GGGCCATCATCGGAGAAAGAGGGGACACATCAACACAGCTGGGAGATTGGCCGAGACC 1317
 Db 421 GlyAlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgPheAlaGlyAla 440
 QY 1318 TCTATCAAGATTGGCTCCCTGGAGAGCCAGAGCTCAACCGAAAGATGTGATATCAC 1377
 Db 441 SerIleLysIleAlaProAlaAspGlyIleAspAlaLysGlnArgMetValIleIleSer 460
 QY 1378 GGGCCACCGGAAGCCCGCTTCAAGCCGAGGAGCGGATCTTGGAACTGAAAGAGAA 1437
 Db 461 GlyProProGluAlaGlnPheLysAlaGlnGlyArgGlyIlePheGlyLysLeuLysGluGln 480
 QY 1438 AACTCTTTAAACCCCAAGAAAGATGAAAGCTGGAAGCGCATATCAGATGCTCTTCC 1497
 Db 481 AsnPhePheGlyProLysGluGluValLysLeuGluAlaHisIleLysValProSerPhe 500
 QY 1498 ACAGCTGCGCGGTGATTGCGAAAGGTGCAAGCCGTGAACGAAGTGCAGAACTTAACC 1557
 Db 501 AlaAlaGlyArgValIleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuThr 520
 QY 1558 AGTGAGAAAGTCATCGTGCCTCGTGAACCAACGCGCAGATGAAATGAGGAATGATGCTC 1617
 Db 521 SerAlaGlnValAlaValaProArgAspGlnThrProAspGluAsnAspGlnAlaValaVal 540
 QY 1618 AGAATTATCGGCACTTCTTGTCTAGCCAGACTGCAACGCGCAAGATCGAGAAATTGTA 1677
 Db 541 LysIleThrGlnGlyHisPheTyrAlaSerGlnLeuAlaGlnArgLysIleGlnGluIleIle 560
 QY 1678 CAACAGGTGAAGCAGCAGAGAGCAAAATACCT 1710
 Db 561 SerGlnValArgArgGlnGlnInProLysPro 571
 RESULT 12
 Q8C2J9 PRELIMINARY; PRT; 579 AA.
 AC Q8C2J9;
 DT 01-MAR-2003 (TRMBLrel. 23, Created)
 DT 01-MAR-2003 (TRMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TRMBLrel. 25, Last annotation update)
 DE Insulin-like growth factor 2.
 GN IGF2BP3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=MOD; TISSUE=Thymus;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium.
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 DR EMBL; AK08465; BAC40370.1; -.
 DR MGI; MGI:1890359; IGF2bp3.
 DR GO; GO:0003676; F.nucleic acid binding; IFA.
 DR InterPro; IPR004087; KH_dom.
 DR InterPro; IPR004088; KH_type_1.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00013; KH; 4.
 DR Pfam; PF00076; rrm; 2.
 DR SMART; SM00322; KH; 4.
 DR SMART; SM00360; RRM; 2.
 DR PROSITE; PSS0084; KH_type_1; 4.

DR PROSITE: PSS0102; RM: 2;
 SQ SEQUENCE 579 AA; 63551 MW; 937E601A95D06B77 CRC64;

Alignment Scores:

Pred. No.: 3, 47e-129 Length: 579
 Score: 1875.00 Matches: 372
 Percent Similarity: 78.22% Conservative: 77
 Best Local Similarity: 64.81% Mismatches: 87
 Query Match: 31.86% Indels: 38
 DB: 11 Gaps: 6

US-09-270-437d-8 (1-3283) x Q8C209 (1-579)

QY 73 ATGAACAAGCTTTATCATCGGGAACCTGAGCCCGCCCTGACCGCGGACGACCTCCGAGC 132
 Db MetanulysleuTyrlleGlyAsnleuSeraphisalaGlyProAlaAspLeuGlySer 20
 QY 133 CTCCTTTGGGACAGGAAGCTGCGCCGCGGACAGGCTCTGCTGAAGTCCGCTACGCC 192
 Db ValPheTyAspAlaTyrlleProValAlaGlyProPheLeuValTyrlleGlyTyrlle 40
 QY 193 TTGCGAGCTACCCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 252
 Db PheValAspCysProAspGlyGlyTyrlleAlaLeuTyrlleGlyAlaLeuSerGlyLys 60
 QY 253 GTGGAATTCATGAGGAATATCATGAGGATGATTACTGATCTGTAAGTAAAGCTAAGAGC 312
 Db MetGluLeuHisGlyLysProMetGluValGluHisSerValProLysAlaGlyAlaGly 80
 QY 313 AGGAATTCATGATTCGAAACATCCCTCTGACCTGACGAGTGAAGGCTGATGACTT 372
 Db ArgLysLeuGlnIleArgAsnIleProProHisLeuGlnIleProGluValLeuAspSerLeu 100
 QY 373 TTGCTCATATGAGGACAGGAGATGGAATGGAACATGACACAGACAGACAGACGACGAC 432
 Db LeuValGlnIleTyrlleValAlaGlySerCysGlnIleValAlaSerHisAspSerGluThrAla 120
 QY 433 GTTGTCAAGCTCATATGACACAGACAGACAGACAGACAGACAGACAGACAGACAGAC 492
 Db ValValAsnValThrTyrlleSerSerLysAspGlnAlaArgGlnAlaLeuAspLysLeuAsn 140
 QY 493 GGGCATATGATTGAAACTACTCTTCAAGATTTCTTCAATCCCGGATGAAAGAGTGAAGC 552
 Db GlyPheGlnIleGlnIleAsnPheThrLeuValAlaIleTyrlleProAspGluThrAlaAla 160
 QY 553 -----TCCGCTTGCGCCCTCAGCGGACCGGAGTGGGAC-----CACTCTTCC 597
 Db GlnGlnAsnProSerProGlnLeuArgGlyArgArgGlyProGlyGlnArgGlySerSer 180
 QY 598 CGGAGACAGGCGCCGCTGCGGAGCACTTCTCAGGCGCAGACAGATTTGATTCGCGCGC 657
 Db ArgGlnAla-----SerProGlySerValSerLysGlnLysProCysAspLeuProLeu 198
 QY 658 CGGATCTGTCCTCCACCCGAGTTGTTGTCATCATCGGAAAGAGGAGCTTGAACATA 717
 Db ArgLeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGlnIleGlyAlaThrIle 218
 QY 718 AAGAAATCATCTAAGACAGCCAGTCCCGGATGATTCATGAAAGAGAACTCTGGA 777
 Db ArgAsnIleThrLysGlnIleThrLysIleAspValHisArgLysGlnIleAsnThrGly 238
 QY 778 GCTGCAGAGAGCCTGTCACATCCATCCGACCCGAGAGGAGCTTCTGAAGATCCGC 837
 Db AlaAlaGlnIleLysSerIleThrIleLeuSerThrProGlnIleTyrlleAlaAlaCysLys 258
 QY 838 ATGATTTCTGAATCATGAGAGAGGAGGATGATGACCAATAGCCGAGAGATTTCT 897
 Db SerIleLeuGlnIleMetHisLysGlnIleAlaGlnAspIleLysPheThrGlnIlePro 278
 QY 898 CTGGAATCTTGCCACACATGCTGCTGTTGAAAGCTGATTTGAAAGAGAGCAGAAAT 957
 Db LeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGlnIleArgAsn 298

QY 958 TTGAAAAATTGAACATGAAACAGGACCAAGATTAACATCTCTTTGACAGATTG 1017
 Db LeuLysIleGlnIleAspThrAspThrLysIleThrIleSerProLeuGlnIleLeu 318
 QY 1018 AGCATATCAACCCGAGAAAGAACATCACTGTGAAGGACAGGATGAGCCCTGTCCACT 1077
 Db ThrLeuTyrlleProGlnIleThrIleThrValLysGlySerValGlnIleThrLysAlaLys 338
 QY 1078 GCTGAGATGAGATTATGAAAGAGCTGCGTGAAGGCTTTGAAATGATGATGCTGCTGT 1137
 Db AlaGlnIleGlnIleMetLysLysIleArgGlnIleSerTyrlleAsnAspIleAlaSerMet 358
 QY 1138 AACACCCAC----- 1146
 Db AsnLeuGlnAlaAsnLeuIleProGlyLeuAsnLeuAlaLeuGlyLeuPheProPro 378
 QY 1147 -----TCCGATTAATCTTCACCTGACCTGTACCCCATACCAAG 1182
 Db ThrSerGlyMetProProProThrSerGlyProProSerThrLeuThrPro----- 395
 QY 1183 TTGACCCGCTCCCGCATCATCACTTTATCCAGACGAGATTTGATCTTTGATC 1242
 Db ProTyrlleProGlnPheGlnIle-----SerGlnIleThrValHisLeuPheIle 412
 QY 1243 CCAACCCAGGCTGTGGGCGCATCATGCGGAAAGAGGCGGACACATCAACAGCTGAGC 1302
 Db ProAlaLeuSerValGlyAlaIleIleGlyLysGlnIleGlnIleLysGlnIleLeuSer 432
 QY 1303 AGATTGCGCGGAGCTTCATCAAGATTTGCGCCGCGGAAAGCCAGACGTCAGCGAAAG 1362
 Db ArgPheAlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLysValArg 452
 QY 1363 ATGCTCATATCAACCGGCGCCACCGAAGCCAGTTCAGGCGCCGAGGAGGATCTTTGGG 1422
 Db MetValIleIleThrGlyProProGlnIleAlaGlnPheLysAlaGlnIleGlyTyrlle 472
 QY 1423 AAAGTGAAGAGGAAATCTTTAACCCCAAGAGAGAGTGAAGCTGAAAGCCATATC 1482
 Db LysIleLysGlnIleGlnIleAsnPheValSerProLysGlnIleValLysLeuGlnAlaHisIle 492
 QY 1483 AGAGTGGCTCTTCCACAGCTGCGGCGGATGTTGGCAAGGTGGCAAGCCGTGAACGAA 1542
 Db ArgValProSerPheAlaAlaGlyArgValIleGlyLysGlyLysIleThrValAsnGln 512
 QY 1543 CTGCAAGAACTTAACAGTGCAGAGATCATGCTGCTGTCGACCAAGCCGAGATGAAT 1602
 Db LeuGlnIleSerLeuSerSerAlaGlnValAlaValProArgAspGlnIleProAspGlnAsn 532
 QY 1603 GAGAAAGTATGCTCAAGATTTATCGGCACTTCTTGTGTAAGCCAGACTGCAAGCGCAG 1662
 Db AspGlnValAlaValLysIleThrGlnIleHisPheTyrlleAlaCysGlnIleAlaGlnArgLys 552
 QY 1663 ATCAGGGAATTTATACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1704
 Db IleGlnIleLeuThrGlnIleValLysGlnIleHisGlnIleGlnIle 566

RESULT 13
 Q9CPN8 PRELIMINARY; PRT; 579 AA.
 AC Q9CPN8;
 ID 01-UTN-2001 (TEMBLrel. 17, last sequence update)
 DT 01-UTN-2001 (TEMBLrel. 17, last sequence update)
 DT 01-OCT-2003 (TEMBLrel. 25, last annotation update)
 DE 10 days embryo cDNA, RIKEN full-length enriched library,
 clone:2610036B18, full insert sequence (igf2 mRNA-binding protein 3)
 DE (Insulin)-like growth factor 2, binding protein 3).
 GN IGF2BP3 OR 261010N1KIK OR MIMP3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Embryo;
 RX MEDLINE=21085660; Pubmed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Atzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
 RA Knehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guerinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Matzarelli U., Mombers P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszewski-Boris A., Yoshida K., Haegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection.",
 RL Nature 409:685-690(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Mori H., Sakakibara S., Imai T., Nakamura Y., Iijima T., Suzuki A.,
 RA Yusa Y., Takeda M., Okano H.,
 RT "Expression of mouse 1972 mRNA-binding protein 3 and its implications
 for the developing central nervous system.",
 RL J. Neurosci. Res. 0:0-0(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Brain, and Olfactory epithelium;
 RX MEDLINE=22388257; Pubmed=12477932;
 RA Strussberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Schemm C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares W.B., Bonaldo M.F., Casavant T.L., Scheltz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullihy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gurnatone P.H.,
 RA Villalón D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,
 RA Fahy U., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
 RA Krzywinski M.I., Skalski U., Smailus D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Maiz M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.",
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Brain, and Olfactory epithelium;
 RX Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 4 KH DOMAINS.
 DR EMBL: AK011689; BAB27779.1; -;
 DR EMBL: AB046173; BAB19755.1; -;
 DR EMBL: BC045138; AAH45138.1; -;
 DR EMBL: BC049082; AAH49082.1; -;
 DR MGI: 1890359; Igt2b3.
 DR GO: 0003676; F: nucleic acid binding; IEA.
 DR InterPro: IPR004087; KH dom.
 DR InterPro: IPR004088; KH type_1.
 DR InterPro: IPR000504; RNA_rec_mot.
 DR Pfam: PF00013; KH; 4.
 DR Pfam: PF00076; Irm; 2.
 DR SMART: SM00322; KH; 4.
 DR SMART: SM00360; RRM; 2.
 DR PROSITE: PSS0084; KH_type_1; 4.

DR PROSITE: PSS0102; RRM; 2.
 SQ SEQUENCE 579 AA; 63574 MW; CABD9A4355B392B7 CRC64;
 Alignment Scores:
 Pred. No.: 3,47e-129 Length: 579
 Score: 1875.00 Matches: 372
 Percent Similarity: 78.22% Conservative: 77
 Best Local Similarity: 64.81% Mismatches: 87
 Query Match: 31.86% Indels: 38
 DB: 11 Gaps: 6
 US-09-270-437d-8 (1-3283) x Q9C9E8 (1-579)
 QY 73 ATGAACAAGCTTTACATCGGGAGACCTTGAGCCCGCCGACCGGACGACCTTCGGGAGC 132
 Db 1 MetanemysleuYrllleglYAsnleuSerAspHisalaglYProAlaAspIeuSer 20
 QY 133 CTCTTTGGGAGCAGAGAAAGCTCCCGCGGGAGCAGGCTGCTGAATCCGGCTACGCC 192
 Db 21 ValPheYAspAlaYsllleProValAlaglYProPheleuValYsThrglYrAla 40
 QY 193 TTGGTGACCTACCCCGACCAAGACTGGCGCCATCCGCCATCGAGACCCCTTCGGGTAA 252
 Db 41 PheValAspCyProAspIeuGlYTriPalaIeuYsAlaIleGlAlaIeuSerGlYs 60
 QY 253 GTGGAATTCATGCGGAAATCATGGAAGTTGATTCTAGTTCTTAAAGCTAAGAGAC 312
 Db 61 MetGluIeuHisglYlYsPProMetGlValGlHisSerValProLYsArgGlAlnArgl 80
 QY 313 AGGAAATTCATGATTCGGAAGAACATCCCTCCTCACCGCGAGTGGAGAGGTGGATGACTT 372
 Db 81 ArgYsIeuGlnIleuYAsnIleProProHisIeuGlnIleuValIleuAspIeu 100
 QY 373 TTGGCTCAATATGGAGCAGTGGAGAAATGGAACAGTCAACAGACAGACAGAAACCGCC 432
 Db 101 LeuValGlnIYrGlYValValGlIeuSerCySglIeuValIeuAsnThrAspSerGlunhRAla 120
 QY 433 GTTGTCACATGTCATATGATGCAACAAGAGAGAGCAAAATAGCCATGAGAGAGTAAAC 492
 Db 121 ValValAsnValThrTySerSerYsAspGlnAlaArgGlnAlaIeuAspYsIeuAsn 140
 QY 493 GGGCATCTGTTGAGAACACTACCTTCAAGATTTCCTCATCCCGATGAAGAGGTAGC 552
 Db 141 GlyPheGlnIeuGlnAsnIeuThrIeuYsValAlaIYrIleProAspIunhRAlaIa 160
 QY 553 -----TCCCTTCGCCCCCTTCAGAGCCAGCCGTGGGAC-----CACTCTTC 597
 Db 161 GlnGlnAsnProSerProGlnIeuArgYsArgYsGlYProGlYGlAlnArgYsSer 180
 QY 598 CGGAGAGCAAGCCAGCCCTTGGGGGACCTTCTCAGGCCAGACAGATTGATTCGCCGCG 657
 Db 181 ArgGlnAla-----SerProGlySerValSerYsGlnIYsProCyAspIeuProIeu 198
 QY 658 CGGATCTCGTCCCGACCCAGTTGTTGTCATCATCGGAAAGAGGAGCTTGACCAT 717
 Db 199 ArgIeuIeuValProThrGlnPheValGlYAlaIleIleGlYsIeuGlYAlaIle 218
 QY 718 AAGAAATCATCTAAGAGAGCCAGTCCCGGATGATATCATAGAAAGAAACTCTGA 777
 Db 219 ArgAsnIleThrlYsGlnIeuIeuSerYsIleAspValHisArgYsGlnIeuThrlY 238
 QY 778 GCTGCAAGAGAGCTTGTACCATTCATGCCACCCCGAGAGGAGCTTCTGAAGATCGCG 837
 Db 239 AlaAlaGluYsSerIleThrlleuSerThProGlnIYrThSerAlaIaIaCYsYs 258
 QY 838 ATGATCTTGAATCATCTAAGAGAGGAGCATGAGCAAACTAGCCGAGAGATTCCT 897
 Db 259 SerIleIeuGlnIleThrlYsGlnIeuValAlaGlnAspIleYsTherIeuThrlYlePro 278
 QY 898 CTGAAATCTGGACACATGAGCTTGTGAGAGACTGATGGAAAGAGAGCAAAAT 957
 Db 279 LeuYsIleIeuAlaHisAsnAsnIeuValGlYArgIeuIleGlYsGlnIYsArgAsn 298

```
QY 958 TTGAAGAAATTTGAACATGAAACAGGACCAAGATTAACATCTCATCTTTCGAGATTGG 1017
DB 299 LeuLysIleGlnGlnAspThrAspThrLysIleThrIleSerProLeuGlnGlnLeu 318
QY 1018 AGCATATACACCCGGAAGAACATCACTGTGAAGGCGACAGTTAGCGCTGTCCAGT 1077
DB 319 ThrLeuTYAspProGlnValThrIleThrValLysGlySerValGlnThrCysAlaLys 338
QY 1078 GCTGAGATAGAGATTATGAGAGAGTGGTGGCTTGAAGATATGCTGGCTGT 1137
DB 339 AlaGlnGlnGlnIleMetLysLysIleArgGlnSerTYGlnAsnAspIleAlaSerMet 358
QY 1138 AACACCCAC----- 1146
DB 359 AsnLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProPro 378
QY 1147 -----TCCGAGATCTTCCAGCTGTACCCCATCAGCAG 1182
DB 379 ThrSerGlyMetProProProThrSerGlyProProSerThrLeuThrPro----- 395
QY 1183 TTTGGCCCGTCCCGCATCATCCTTATCCAGAGCGAGATTGTGAATCTTTCATC 1242
DB 396 -----ProTYrProGlnPheGlnGln---SerGlnThrGlnThrValHisLeuPheIle 412
QY 1243 CCAACCCAGGCTGTGGGCGCCATCATCGGGAAGAGGGGCGACACATCAACAGCTGGCG 1302
DB 413 ProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlnHisIleLysGlnLeuSer 432
QY 1303 AGATTGCGCGAGCCTCTATCAAGATTGCCCTGCGGAAGCGCAGAGTCAGCGGAAG 1362
DB 433 ArgPheAlaGlyAlaSerIleLysIleAlaProAlaGlnAlaProAspAlaLysValArg 452
QY 1363 ATGTCATCATCACCAGGCGCCACCGGAAGCCAGTTCAAGGCCCGAGGACGGATCTTTGGG 1422
DB 453 MetValIleIleThrGlyProProGlnAlaGlnPheLysAlaGlnGlyArgIleTYrGly 472
QY 1423 AAACGGAAGAGAGAACTTCTTTAACCCCAAGAGAGAGAGAGAGCTGGAAGCGCATATC 1482
DB 473 LysIleLysGlnGlnAsnPheValSerProLysGlnGlnValLysLeuGlnAlaHisIle 492
QY 1483 AGAGTCCCTCTTCCACAGCTGCGCGGATGTGGCAAGGTGGCAAGCCGTGACGAA 1542
DB 493 ArgValProSerPheAlaAlaGlyArgValIleGlyLysGlnGlyLysThrValAsnGln 512
QY 1543 CTGCAAGAACTTAACCAAGTGCAGAAATCTGCTGCTGTAACCAAGCCAGATGAAT 1602
DB 513 LeuGlnSerLeuSerSerAlaGlnValValAlaProArgAspGlnThrProAspGlnAsn 532
QY 1603 GAGGAAGTATGATCGTCAAGATTAATGCGGCACTTCTTGTAGCCAGACTGCAGCGGCAAG 1662
DB 533 AspGlnValValValLysIleThrGlyHisPheTYrAlaCysGlnValAlaGlnArgLys 552
QY 1663 ATCAGGGAATTTGACAAACAGTGAAGCAGAGAGAGAGAA 1704
DB 553 IleGlnGlnIleLeuThrGlnValLysGlnHisGlnGlnGln 566

RESULT 14
000425 PRELIMINARY; PRT, 579 AA.
AC 000425;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative RNA binding protein KOC (KOC).
GN KOC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Mueller-Pillasch F., Lacher U., Wallrapp C., Et Al.;
RL Oncogene 0:0-0(0).
```

```
RN [2]
RP SEQUENCE FROM N.A.
RC TISUB=Pancreas;
RA Mueller-Pillasch F., Lacher U., Wallrapp C.;
RL Submitted (0CF-1996) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR EMBL; U97188; AAC35208.1; -.
DR EMBL; U76705; AAD09223.1; -.
DR GO; GO:0005737; C:cytoplasm; TAS.
DR GO; GO:0003723; F:RNA binding; TAS.
DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR GO; GO:0006412; P:protein biosynthesis; TAS.
DR GO; GO:0006396; P:RNA processing; TAS.
DR InterPro; IPR004087; KH_dom.
DR InterPro; IPR004088; KH_type_1.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00013; KH; 4.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00322; KH; 4.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS50084; KH_type_1; 4.
DR PROSITE; PS50102; RRM; 2.
DR PROSITE; PS0030; RRM_RNP_1; FALSE NEG.
SQ SEQUENCE 579 AA; 63720 MW; AE5C3A8EE3C135C5 CRC64;
```

Alignment Scores:

Pred. No.:	1,14e-128	Length:	579
Score:	1868.00	Matches:	368
Percent Similarity:	78.52%	Conservative:	78
Best Local Similarity:	64.79%	Mismatches:	96
Query Match:	31.74%	Indels:	26
DB:	4	Gaps:	6

US-09-270-437D-8 (1-3283) x 000425 (1-579)

```
QY 73 ATGAACAAGCTTTTCATCGGGAACCTGAGCCCGCCGTCACCGCGAGACTCCGCGAG 132
DB 1 MetAsnLysLeuTYrIleGlyAsnLeuSerIleAsnAlaAlaProSerAspLeuGlnSer 20
QY 133 CTCCTTGGGCGACAGAAAGCTGCCCCGCGGACAGAGTCTGTAAGTCCGCGTACGCC 192
DB 21 IlePheLysAspAlaLysIleProValSerGlyProPheLeuValLysThrGlyTYrAla 40
QY 193 TTCGTGACTACCCCGACAGAACTGGCGCATTCGCGCATGAGACCTCTCGGGTAA 252
DB 41 PheValAspCysProAspGlnSerTrpAlaLeuLysAlaIleGlnAlaLeuSerGlyLys 60
QY 253 GTGGAATTGCAATGGGAATATCATGAGTGTATCTACAGTCTTAATAAAGCTAAGAGC 312
DB 61 IleGlnLeuHisGlyLysProIleGlnValGlnHisSerValProLysArgGlnArgIle 80
QY 313 AGGAATAATTCAGATTCGAAACATCCCTCCATCGAGTGGAGGTGGATGAGACTT 372
DB 81 ArgLysLeuGlnIleAlaGlnIleProProHisLeuGlnTrpGlnValLeuAspSerLeu 100
QY 373 TTGCTCAATATGGGACAAGTGAAGATGTGAACAGTCAACACAGACACAGAAACGCC 432
DB 101 LeuValGlnTYrGlyValValGlnSerCysGlnGlnValAsnThrAspSerGlyThrAla 120
QY 433 GTTGTCACAGCTCACTATGCAACAGAGAAACCAAAAAATAGCATGAGAGACTAAGC 492
DB 121 ValValAsnValThrTYrSerSerLysAspGlnAlaArgGlnAlaLeuAspLysLeuAsn 140
QY 493 GGGCATCAGTTTGAAGACTACTCTTCAAGATTTCTTACATCCCGGATGAAGAGGTGAGC 552
DB 141 GlyPheGlnLeuGlnAsnPheThrLeuLysValAlaIleTYrIleProAspGlnMetAlaIle 160
QY 553 TCCCTTCGCCCTCAG-----CGAGCCAGCGT-----GGGAGCAACTCTTCCGG 600
DB 161 GlnGlnAsnProLeuGlnGlnProArgIlyArgArgGlyLeuGlyGlnArgGlySerSer 180
QY 601 GAGCAAGCGCACCGCCCTGGGGGAGCATCTTCAGAGCGCAGACAGATTGATTTCCGCGTGGG 660
```


This Page Blank (uspto)